

th J, Miller RH, Purcell RH;
1; 96-139709/14.
25DB; T16650.
and amino acid sequence of HCV envelope 1 and core proteins -
to determine HCV genotype and as vaccines against HCV infection
im 4; Page 214; 340pp; English.
936-R92987 are HCV core proteins derived from 52 different HCV
plates. Isolated cDNA sequences are used for the prodn. of primers
ful for detecting the presence of HCV in a sample, the primers
also useful for HCV genotyping. Proteins encoded by the cDNAs
be used in vaccines for immunising against HCV infection. The
teins may also be used to detect antibodies against HCV in serum,
iva, lymphocytes or other mononuclear cells. The antibodies may
used in the prevention of HCV infection.
ence 191 AA;

Match 99.5%; Score 777; DB 17; Length 191;
ocal Similarity 90.8%; Pred. No. 2.12e-61;
s 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
1 matnbpqrktkntnrrrrpXXXXPFGGQIVGGVYLLPRGRGXGVRATKTSERSQPRG 60
|||||
1 MSTNPKPQRKTNTNRRPXXXXPFGGQIVGGVYLLPRGRGXGVRATKTSERSQPRG 60
1 rrpdpkarrsegrswagqypwplynecgagwvllprgrsrpswq 109
|||||
1 RRQPIPKAXRXEGRSWAQGYPMPLYGNECCGMAXWLLSPRGSRPNWGP 109
|||||

4
373 standard; Protein; 191 AA.
ACT-1996 (first entry)
itis C virus isolate 28 core protein.
E; envelope 1; core protein; HCV genotyping; antibody; vaccine;
itis C virus.
-05315-AZ.
EB-1996.
UG-1995; U10398.
UG-1994; US-290665.
H) US DEPT HEALTH & HUMAN SERVICES.
H) US SEC DEPT HEALTH.
J, Miller RH, Purcell RH;
96-139709/14.
DB; T16647.
and amino acid sequence of HCV envelope 1 and core proteins -
to determine HCV genotype and as vaccines against HCV infection
m 4; Page 211-212; 340pp; English.
35-R92987 are HCV core proteins derived from 52 different HCV
aces. Isolated cDNA sequences are used for the prodn. of primers
ul for detecting the presence of HCV in a sample, the primers
also useful for HCV genotyping. Proteins encoded by the cDNAs
ce used in vaccines for immunising against HCV infection. The
ins may also be used to detect antibodies against HCV in serum,
va, lymphocytes or other mononuclear cells. The antibodies may
ed in the prevention of HCV infection.
nce 191 AA;

atch 99.5%; Score 777; DB 17; Length 191;
al Similarity 90.8%; Pred. No. 2.12e-61;
s 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
matnbpqrktkntnrrrrpmdvdfpgggqivggvylprgrprlgvratktsersqrg 60

Oy 1 MSTNPKPQRKTNTNRRPXXXXPFGGQIVGGVYLLPRGRGXGVRATKTSERSQPRG 60
Db 61 rrpdpkarrsegrswagqypwplynecgagwvllprgrsrpswq 109
Oy 61 RRQPIPKAXRXEGRSWAQGYPMPLYGNECCGMAXWLLSPRGSRPNWGP 109

RESULT 5
ID R67591 standard; Protein; 502 AA.
AC R67591;
DT 07-SEP-1995 (first entry)
DE Hepatitis C virus SR037-5' gene product.
KW Hepatitis C virus; HCV; non-A non-B; SR037-5'; treatment.
OS Hepatitis C virus.
PN J06319563-A.
PD 22-NOV-1994.
PF 13-MAY-1993; 147133.
PR 13-MAY-1993; JP-147133.
PA (IMMO) IMMUNO JAPAN KK.
DR WPI; 95-040318/06.
DR N-PSDB; 079143.
PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the
treatment of hepatitis C
PS Claim 21; Page 36; 41pp; Japanese.
CC 079143 is the hepatitis C virus (HCV) gene SR037-5' cDNA, it encodes
the protein described in R67591. Both the cDNA and protein can be
used in the treatment of HCV infection
SQ Sequence 502 AA;

Query Match 99.5%; Score 777; DB 12; Length 502;
Best Local Similarity 90.8%; Pred. No. 2.12e-61;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Db 1 matnbpqrktkntnrrrrpmdvdfpgggqivggvylprgrprlgvratktsersqrg 60
Oy 1 MSTNPKPQRKTNTNRRPXXXXPFGGQIVGGVYLLPRGRGXGVRATKTSERSQPRG 60
Db 61 rrpdpkarrsegrswagqypwplynecgagwvllprgrsrpswq 109
Oy 61 RRQPIPKAXRXEGRSWAQGYPMPLYGNECCGMAXWLLSPRGSRPNWGP 109

RESULT 6
ID R67589 standard; Protein; 502 AA.
AC R67589;
DT 07-SEP-1995 (first entry)
DE Hepatitis C virus YS117-5' gene product.
KW Hepatitis C virus; HCV; non-A non-B; YS117-5'; treatment.
OS Hepatitis C virus.
PN J06319563-A.
PD 22-NOV-1994.
PF 13-MAY-1993; 147133.
PR 13-MAY-1993; JP-147133.
PA (IMMO) IMMUNO JAPAN KK.
DR WPI; 95-040318/06.
DR N-PSDB; 079141.
PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the
treatment of hepatitis C
PS Claim 19; Page 34; 41pp; Japanese.
CC 079141 is the hepatitis C virus (HCV) gene YS117-3' cDNA, it encodes
the protein described in R67588. Both the cDNA and protein can be
used in the treatment of HCV infection
SQ Sequence 502 AA;

RESULT	1	
ID	A50346	standard; DNA; UNC; 327 BP.
AC	A50346;	
VI	ei039185	
DT	07-MAR-1997	(Rel. 51, Created)
DT	07-MAR-1997	(Rel. 51, Last updated, Version 1)
DE	Sequence 1 from Patent WO9613590.	
WE	.	
WM	unidentified	
VS	3	
VC	unclassified.	
RI	[1]	
RP	1-327	
RA	Maertens G., Stuyver L.;	
RT	"NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE	
TT	AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS";	
LL	Patent number WO9613590-A1, 09-MAY-1996.	
CC	INNOCENTIS NV (BE).	
CC	Other publication AU 3844095 960523	
HH	Key	Location/Qualifiers
HH		
HH	source	1..327
TT		/organism="unidentified"
TT	Sequence 327 BP; 56 A; 102 G; 104 G; 51 T; 14 other;	

[illegible]

RESULT	2	HCVN3	506 bp	DNA	VRL	29-APR-1994
OCUS		Hepatitis C virus core gene, HN3 isolate.				
DEFINITION		X76410				
ACCESSION		U01497				
KEYWORDS		core; core protein.				
SOURCE		Hepatitis C virus.				
ORGANISM		Hepatitis C virus.				
		Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis C-like viruses.				

REFERENCE	1 (bases 1 to 506)
AUTHORS	Qu,D.
TITLE	Direct Submission
JOURNAL	Submitted (22-NOV-1993) to the EMBL/GenBank/DBJ databases. D. Qu, INSERM Unite 271, U. de Recherche sur les Hepatites, le Sida et les Retrovirus Humains, 151 cours Albert Thomas, F 69424 Lyon Cedex 03, FRANCE
REFERENCE	2 (bases 1 to 506)
AUTHORS	Qu,D., Hantz,O., Gouy,M., Vitvitski,L., Li,J.S., Berby,F., Tong,S.P. and Trepo,C.
TITLE	Heterogeneity of hepatitis C virus genotypes in France
JOURNAL	J. Gen. Virol. 75 (Pt 5), 1063-1070 (1994)

FEATURES	
source	Location/Qualifiers
	1..506
	/organism="Hepatitis C virus"
	/isolate="HN3"
misc_feature	29..506
	/note="region corresponding to the core protein of HCV"
BASE COUNT	86 a 166 c 166 g 88 t
ORIGIN	

Query Match	91.8%	Score 291	DB 117	Length 506
Best Local Similarity	91.7%	Pred. No. 4,45e-221		
Matches 300	Conservative	4	Mismatches 23	Indels 0
Db	29	atgagcaagatcctaaacctcaagaagaacacaaactgaactcaacccgcgcacag	88	
Qy	1	ATGAGCAGCAATCTTAAACTCTAAAGAAACCAACCTAAGACACACGCGCGCCTCAK	60	
Db	89	gacgtcaagttcccgcggtggtgcagatcgttggtagttacctgttgcgcgcaag	148	
Qy	61	CGSGTNNNNNNCCGGTGGCGGTGAGATCTGTGTGGAGCTTTACTCTTGC	120	
Db	149	ggcccccaagttgggtgtgcgcgcgactaggaaactccgagcaggtcgcaacctcgtggc	208	
Qy	121	GGCCCGACGNNGGGTGTGGCGCGCACTAGGACACTCCGACGGGTCAACACCTCTGTGGC	180	
Db	209	aggcgacagcactatcccgaaggtcgtcgcgcgcaggcaggtcctgggctcagccggg	268	
Qy	181	AGCCGACAGCCTATCCCAAGGCTCGCGYCCGAGGSCAGCTCTGGCTCAGCCCGGG	240	
Db	269	taaccttggccctctatagcaacagaggttggcgggtgggcagggtggctcgtccccc	328	
Qy	241	TATCCTTGGCCCCCTCTATGGCAATGAGGCTTGGGTGGCGGATGGCTCTCTCCCC	300	
Db	329	cgtggcctctcggcctagttggggcccc	355	
Qy	301	CGCGGCTCTCGGCCAATTTGGGGCCCC	327	

3	HCU10204	573 bp	RNA	VRL	26-AUG-1994
DEFINITION	Hepatitis C virus isolate P10 core protein gene.....				
ACCESSION	U10204				
NID	9532392				
KEYWORDS	.				
SOURCE	Hepatitis C virus.				
ORGANISM	Hepatitis C virus				
REFERENCE	Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis G-like viruses.				
AUTHORS	Buhh, J., Purcell, R.H. and Miller, R.H.				

```

RESULT 1
ID ID A50346 standard; DNA; UNC; 327 BP.
AC AC A50346;
WI el039185
OT 07-MAR-1997 (Rel. 51, Created)
DT 07-MAR-1997 (Rel. 51, Last updated, Version 1)
DE Sequence 1 from Patent W09613590.
KW .
   unidentified
   unclassified.
   [1]
   [1]
   1-327
   1-327
   Maertens G., Stuyver L.;
   "NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE
   AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS";
   Patent number W09613590-A/1, 09-MAY-1996.
   INNOGENETICS NV (BE).
   Other publication AU 3844095 960523
   Key Location/Qualifiers
   source 1..327
   /organism="unidentified"
   Q Sequence 327 BP; 56 A; 102 C; 104 G; 51 T; 14 other;
   Query Match 100.0%; Score 317; DB 143; Length 327;
   Best Local Similarity 100.0%; Pred. No. 7.68e-244;
   Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

bb 1 atgagcagcaatcctaaacctcaagaataacaaacgtaacacacccgcgcctcak 60
yy 1 ATGCAGCAGCAATCCTTAAACCTCAAGAAATAACCAAGCTAAACACCAACCGCGCCTCAK 60

bb 61 ggggtnnnnnnccgggtggcggtcagatcggtgtgtagttacgtgttgccgcgagg 120
yy 61 GGGGTNNNNNNCCGGGTGGCGGTCAGATCGTGTGGTGGAGTTACGTGTGGCGCGCAGG 120

b 121 ggcacagcagnggggtgtgcgcgagctaggaagacttcgagcggtcacaaacctgtggc 180
yy 121 GCGCCACAGNNGGVTGTGCGCGGCACTAGGAAGACTTCCGACGGGTCAAACTCGTGGC 180

b 181 aggcagacagcctatccccaaaggtcgcgcgcagcggcgaggtcctggcctcagcccg 240
yy 181 AGCGCAGACGCCTATCCCCAAGGCTCGCGGCGCAGGAGGTCTCGGCTAGCGCCGGG 240

b 241 tatccttgccctctatggcaatgagcgtcggggtgggcgggntggctcctgtcccc 300
yy 241 TATCTCTGCCCTCTATGGCAATGAGGCTCGGGTCGAGGGCAGGTCTCGGCTAGCGCCGG 300

b 301 cgcgctctcgcccaattggggccc 327
yy 301 CGCGCTCTCGGCCAATTGGGGCCCC 327

```

RESULT	2
TITLE	HCV/HN3
AUTHOR	HCYNH3
JOURNAL	Hepatitis C virus core gene, HN3 isolate.
PAGE	506 bp
KEYWORDS	VRL
SOURCE	Hepatitis C virus core gene, HN3 isolate.
ORGANISM	X76410
IDENTIFICATION NO.	.943497
SYNOPSIS	ID .943497
ABSTRACT	core; core protein.
REFERENCES	Hepatitis C virus.
CROSS-REFERENCE	Hepatitis C virus
NOTES	Hepatitis C virus
COMMENTS	Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
REMARKS	Hepatitis C-like viruses.
DATE	29-APR-1994

```

REFERENCE 1 (bases 1 to 506)
AUTHORS Qu, D.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-1993) to the EMBL/GenBank/DDBJ databases. D. Qu,
INSERM Unite 271, U. de Recherche sur les Hepatites,, le Sida et
les Retrovirus Humains, 151 cours Albert Thomas, F 69424 Lyon Cedex
03, FRANCE

REFERENCE 2 (bases 1 to 506)
AUTHORS Qu, D., Hantz, O., Gouy, M., Vitvitski, L., Li, J.S., Berby, F.,
Tong, S.P. and Trepo, C.
TITLE Heterogeneity of hepatitis C virus genotypes in France
J. Gen. Virol. 75 (pt 5), 1063-1070 (1994)
MEDLINE 94231157
FEATURES
source Location/Qualifiers
1..506
/organism="Hepatitis C virus"
/isolate="HN3"
misc_feature 29..506
note="region corresponding to the core protein of the
polyprotein of HCV"
BASE COUNT 86 a 166 c 166 g 88 t
IGIN

```

Query Match	91.8%;	Score 291;	DB 117;	Length 506;	
Best Local Similarity	91.7%;	Pred. No. 4.45e-221;			
Matches 300;	Conservative 4;	Mismatches 23;	Indels 0;	Gaps 0;	
bb	29	atgacgacgaatcctaaacctcaaaaggaaacaaacacgtaacatcaaccgccgccccacag	88		
yy	1	ATGACGACGAATCTTAAACCTCAAAAGAAACAAACCAAGCTAAGCAACACCGCGCGCCTCAK	60		
bb	89	gaagtcgaagttcccgggcggtggtcagatcgttgttggaagtttacctgtgccgcgcaggg	148		
yy	61	GGSGTNNNNNNNCGGGTGGCGGTGAGATCGTTGGTGGAGGTTTACTCTGTCGGCGGCGAGG	120		
bb	149	ggcccagcagtggtgtgcgcgcgcactagaagaactccgcagcggtcgcaacctcgtgcg	208		
yy	121	GGCCCCAGGNNGGTGTGCGCGGCACTAGGAGACATTCGAGGGGTACAAACCTCGTGGC	180		
bb	209	aggcgacagacctatcccaaggctcgtcgccgcagggcaggtcctgggtcagcccggg	268		
yy	181	AGCGCAGACGCTATCCCAAGGCTCGYCGYCGAGGGCAGGTCCTGGGGTCAGCCCGGG	240		
bb	269	tacctttggccctctatggcaacgagggctgggggtggcgaggtggctcctgtcccc	328		
yy	241	TATCCTTGGCCCTCTATGGCAATAGGGGTGCGGGTGGCGGGGNTGCTCCTGTGCCCC	300		
bb	329	cgtggcctcggcctagttggggcccc	355		
yy	301	CGCGGCTCTCGGCCCAATATGGGGCCCC	327		

3
 RESULT
 LOCUS
 DEFINITION
 ACCESSION
 U10204
 ID
 g532392
 KEYWORDS
 Hepatitis C virus.
 ORGANISM
 Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepatitis C-like viruses.
 REFERENCE
 1 (strs)
 AUTHORS
 Bukh J., Purcell R.H. and Miller R.H.

CCCCC 300
23
- 74
9

38/227

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 26
SEQUENCE 191 AA: 20774 MW: 184617 CN:

```

Query Match      ...99.5%; Score 777; DB 13; Length 191;
Best Local Similarity 90.8%; Pred. No. 2.00e-59;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Caps 0;

Db      1 MSTNPKPQKTKRNTNRPPDMVKPGGGQIVGGVYLLPRGPRPLGVRAKTKTSERSQPRG 60
QY      1 MSTNPKPQKTKRNTNRPPDMVKPGGGQIVGGVYLLPRGPRPLGVRAKTKTSERSQPRG 60

Db      61 RROP1PKARRSEGRSQAQGYPMWLYGNECCGACGALLSPGRSRP SWGP 109
QY      61 RROP1PKARRSEGRSQAQGYPMWLYGNECCGACGALLSPGRSRP SWGP 109

```

RESULT	2
ID	PCT-US95-10398-192
XX	STANDARD;
AC	PRT; 191 AA.
XX	xxxxxx
XX	01-JAN-1900
XX	DT
XX	Sequence 192, Application PC/TUS9510398..
XX	Sequence 192, Application PC/TUS9510398
CC	GENERAL INFORMATION:
CC	APPLICANT: BUKH, J., MILLER, R.H. AND
CC	APPLICANT: PURCELL, R.H.
CC	TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC	TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC	TITLE OF INVENTION: CORE GENES OF ISOATES OF HEPATITIS C VIRUS
CC	TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC	TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC	NUMBER OF SEQUENCES: 263

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC
CC INFORMATION FOR SEQ ID NO: 192:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: 28
SQ SEQUENCE 191 AA; 20744 MW; 181593 CN;

Query Match	99.5%	Score 777;	DB 13;	Length 191;
Best Local Similarity	90.8%;	Pred. No. 2.00e-59;		
Matches	99;	Conservative 1;	Mismatches 9;	Indels 0; Gaps 0;

[illegible][illegible]

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ai

5

CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 158:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: SW1
CC SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQKTKRNTNRRPQDKFPGGQIVGGVYLLPRRGLGVRATKTSERSQPRG 60
Qy 1 MSTNPKQKTKRNTNRRPQDKFPGGQIVGGVYLLPRRGLGVRATKTSERSQPRG 60

Db 61 RQPIPKARPERGTWAGPGYPPPLYNGECGAGWALLSPGRSRPWSGP 109
Qy 61 RQPIPKARPERGTWAGPGYPPPLYNGECGAGWALLSPGRSRPWSGP 109

RESULT 4

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ai

6

ID PCT-US95-10398-156 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 156, Application PC/TUS9510398.
XX
CC Sequence 156, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 156:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: US11
CC SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQKTKRNTNRRPQDKFPGGQIVGGVYLLPRRGLGVRATKTSERSQPRG 60

Qy 1 MSTNPKQKTRNTNRPRXXXXXPGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60
|||||
Db 61 RRQIPKARPEGRTPAOPGYPWPLYNGEGCGMACWLLSPRGSRFSWGP 109
|||||
Qy 61 RRQIPKAXRGEGRSWAQPCYPWPPLYNGEGCGMACWLLSPRGSRPNWGP 109
|||||
RESULT 5
ID PCT-US95-10398-157 STANDARD; PRT; 191 AA.
XX XXXXX
XX 01-JAN-1900
DE Sequence 157, Application PC/TUS9510398.
XX
CC Sequence 157, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 157:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:

CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: S14
SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;
Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQKTRNTNRPRQDVKFPFGGQIVGGVYLLPRGPRGLGVRATRKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTRNTNRPRXXXXXPGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60
|||||
Db 61 RRQIPKARPEGRTPAOPGYPWPLYNGEGCGMACWLLSPRGSRFSWGP 109
|||||
Qy 61 RRQIPKAXRGEGRSWAQPCYPWPPLYNGEGCGMACWLLSPRGSRPNWGP 109
|||||
RESULT 6
ID PCT-US95-10398-196 STANDARD; PRT; 191 AA.
XX XXXXX
XX 01-JAN-1900
DE Sequence 196, Application PC/TUS9510398.
XX
CC Sequence 196, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.rai

9

CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 196:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORGANISM: homo sapiens
CC INDIVIDUAL ISOLATE: Z7
SQ SEQUENCE 191 AA; 20788 MW; 180939 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRTNRRPMDVKFPGGQIVGGVYLLPRGRPLGLVTRTKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRTNRRPMDVKFPGGQIVGGVYLLPRGRPLGLVTRTKTSERSQPRG 60
|||||

Db 61 RROQIPKARXEGRSWAQCPWPPLYNCGCGWAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RROQIPKARXEGRSWAQCPWPPLYNCGCGWAGWLLSPRGSRPWSGP 109
|||||

RESULT 7
ID PCT-US95-10398-159 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 159, Application PC/TUS9510398.
XX
CC Sequence 159, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.rai

10

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 159:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homo sapiens
CC INDIVIDUAL ISOLATE: S18
SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRTNRRPMDVKFPGGQIVGGVYLLPRGRPLGLVTRTKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRTNRRPMDVKFPGGQIVGGVYLLPRGRPLGLVTRTKTSERSQPRG 60
|||||

Db 61 RROQIPKARXEGRSWAQCPWPPLYNCGCGWAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RROQIPKARXEGRSWAQCPWPPLYNCGCGWAGWLLSPRGSRPWSGP 109
|||||

RESULT 8
ID PCT-US95-10398-160 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 160, Application PC/TUS9510398.
XX
CC Sequence 160, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993

2

STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07280
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FOREMBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5521.US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 967 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 967 AA; 106397 MW; 5014170 CN;

[illegible]

RECORD NO.	ID	STANDARD;	PRT;
1	PCT-US95-01087-13	967 AA.	

XX
XXXXXX

XX
XXXX
sequence is, application 10/1003331001.

CC sequence is, Application ID: 103330100,
CC GENERAL INFORMATION:

CC APPLICANT: WATANABE, SHINICHI
CC APPLICANT: YAMAGUCHI, JUI-IE

CC	APPLICANT: DESAI, SURESH M.
CC	APPLICANT: DEVARE SUSHIL. G.

CC TITLE OF INVENTION: MAMMALIAN ENVELOPE

CC NUMBER OF SEQUENCES: 22

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: ABBOTT LABORATORIES D311/AP6D
CC CREDIT: ANT PROGRAM D311 D3D

CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/01087
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: PORCMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 967 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 967 AA; 106397 MW; 5014170 CN;

Query Match 99.2%; Score 775; DB 13; Length 967;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQKTRKNTNRNRPQVFGGQIVGGVYLLPRRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTRKNTNRNRPQVFGGQIVGGVYLLPRRGLGVTRATKTSERSQPRG 60
|||||

Db 61 RQPIKARPCGRTWAQPGYMPLYNGEGCGWAGWLLSPRGSRPWSGP 109
|||||

Qy 61 RQPIKARPCGRTWAQPGYMPLYNGEGCGWAGWLLSPRGSRPWSGP 109
|||||

RESULT 11
ID US-08-188-281B-13 STANDARD; PRT; 967 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 13, Application US/08188281B.
XX
CC Sequence 13, Application US/08188281B
CC Patent No. 5610009
CC GENERAL INFORMATION:
CC APPLICANT: WATANABE, SHINICHI
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP60
CC STREET: ONE ABBOTT PARK ROAD

CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/188,281B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: PORCMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 967 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 967 AA; 106397 MW; 5014170 CN;

Query Match 99.2%; Score 775; DB 7; Length 967;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQKTRKNTNRNRPQVFGGQIVGGVYLLPRRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTRKNTNRNRPQVFGGQIVGGVYLLPRRGLGVTRATKTSERSQPRG 60
|||||

Db 61 RQPIKARPCGRTWAQPGYMPLYNGEGCGWAGWLLSPRGSRPWSGP 109
|||||

Qy 61 RQPIKARPCGRTWAQPGYMPLYNGEGCGWAGWLLSPRGSRPWSGP 109
|||||

RESULT 12
ID PCT-US95-01087-12 STANDARD; PRT; 1648 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 12, Application PC/TUS9501087.
XX
CC Sequence 12, Application PC/TUS9501087
CC GENERAL INFORMATION:
CC APPLICANT: WATANABE, SHINICHI
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP60
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK

CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/01087
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: FOREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1648 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1648 AA; 178869 MW; 14581021 CN;

Query Match	99.2%	Score 775	DB 13	Length 1648
Best Local Similarity	89.9%	Pred. No. 3.03e-59		
Matches	98	Conservative	2	Mismatches 9
			Indels	0
			Gaps	0
Db	1	MSTNPKPQKTKNTNRRPDQVKEPGCGQIVGGVYLLPRGRPLGVRATKTSERSQPRG	60	
Qy	1	MSTNPKPQKTKNTNRRPXXXXPGGQIVGGVYLLPRGRPRXGVRATKTSERSQPRG	60	
Db	61	RROP1PKARRPGRWAOQGYPWLYNCGCGGACGALLSPGRSRP SWCP	109	
Qy	61	RROP1PKAXRKGERSWAOQGYPWLYNCGCGGACGALLSPGRSRPNWCP	109	

RESULT	13
ID	PCT-US94-07280-12
XX	STANDARD;
AC	PRF; 1648 AA.
XX	
XX	
DT	01-JAN-1900
XX	
DE	Sequence 12, Application PC/TUS9407280.
XX	
CC	Sequence 12, Application PC/TUS9407280
CC	GENERAL INFORMATION:
CC	APPLICANT: WATANABE, SHINICHI
CC	APPLICANT: YAMAGUCHI, JULIE
CC	APPLICANT: DESAI, SURESH M.
CC	APPLICANT: DEVARE, SUSHIL G.
CC	TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC	TITLE OF INVENTION: ENVELOPE GENES
CC	NUMBER OF SEQUENCES: 22
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
CC	STREET: ONE ABBOTT PARK ROAD
CC	CITY: ABBOTT PARK
CC	STATE: IL

CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/07280
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1648 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1648 AA; 178869 MW; 14581021 CN;

Query Match	99.2%;	Score 775;	DB 12;	Length 1648;
Best Local Similarity	89.9%;	Pred. No. 3.03e-59;		
Matches	98;	Conservative	2;	Mismatches 9; Indels 0; Gaps 0;

Db	1	MSTNPKPQKTKNTNRRPQDVKEP	GGCGIVGGVYLLPRGRPLGVTRTKTTSRQPRG	60
Qy	1	MSTNPKPQKTKNTNRRPQDVKEP	GGCGIVGGVYLLPRGRPLGVTRTKTTSRQPRG	60
Db	61	RROP1PKARRPGRTHAQP	GYPWPLLYNCGEGCGAGWLLSPGRSPNCP	109
Ov	61	RROP1PKARRPGRTHAQP	GYPWPLLYNCGEGCGAGWLLSPGRSPNCP	109

[illegible]

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.rai

17

CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/188,281B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1648 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1648 AA; 178869 MW; 14581021 CN;

Query Match 99.2%; Score 775; DB 7; Length 1648;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGVRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGVRATKTSERSQPRG 60
|||||
Db 61 RRQPIPKARPRGRTWAQPGYWPPLYNGECGCGWALLSPRGSRPNSGP 109
|||||
Qy 61 RRQPIPKARPRGRTWAQPGYWPPLYNGECGCGWALLSPRGSRPNSGP 109
|||||

RESULT 15
ID PCT-US94-07280-1 STANDARD; PRT; 3011 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
XX Sequence 1, Application PC/TUS9407280.
XX
CC Sequence 1, Application PC/TUS9407280
CC GENERAL INFORMATION:
CC APPLICANT: WATANABE, SHINICHI
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC TITLE OF INVENTION: ENVELOPE GENES
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.rai

18

CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/07280
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3011 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 3011 AA; 327108 MW; 47577590 CN;

Query Match 99.2%; Score 775; DB 12; Length 3011;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGVRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGVRATKTSERSQPRG 60
|||||
Db 61 RRQPIPKARPRGRTWAQPGYWPPLYNGECGCGWALLSPRGSRPNSGP 109
|||||
Qy 61 RRQPIPKARPRGRTWAQPGYWPPLYNGECGCGWALLSPRGSRPNSGP 109
|||||

Search completed: Mon Feb 23 11:06:01 1998
Job time : 12 secs.

* * *



(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

protein - protein database search, using Smith-Waterman algorithm

```
Run on: Mon Feb 23 11:03:09 1998; MasPar time 11.95 Seconds
226.191 Million cell updates/sec
Tabular output not generated.
```

0753 00 030 0753 2

Description: (1-109) from US08836075A.ppt

Sequence: 1 MSTNPKPORKTKRNTNRRPX

<
 L
 T
 3
 6
 T
 T
 T
 T
 T

Gap 11

Searched: 243740 cons 24700054 residues

Listing first 45 summaries

Deputy Secretary

1:PCT91 2:PCT92 3:PCT93 4:P

15:U79B 16:U800 17:U801 18:

00:00:00:00 00:00:00:00

36:0820 37:0821 38:0822 39:

50:U834 51:U835 52:U836 53:

37:0841 38:0842 39:0843 00:

71:U60 72:NEWP 73:NEWU6 74:

Statistics: Mean 27.872: Variance 128.07

—

score greater than or equal to the score

SUMMARIES

Result	Query
--------	-------

101 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 1992 1991 1990 1989 1988 1987 1986 1985 1984 1983 1982 1981 1980 1979 1978 1977 1976 1975 1974 1973 1972 1971 1970 1969 1968 1967 1966 1965 1964 1963 1962 1961 1960 1959 1958 1957 1956 1955 1954 1953 1952 1951 1950 1949 1948 1947 1946 1945 1944 1943 1942 1941 1940 1939 1938 1937 1936 1935 1934 1933 1932 1931 1930 1929 1928 1927 1926 1925 1924 1923 1922 1921 1920 1919 1918 1917 1916 1915 1914 1913 1912 1911 1910 1909 1908 1907 1906 1905 1904 1903 1902 1901 1900 1899 1898 1897 1896 1895 1894 1893 1892 1891 1890 1889 1888 1887 1886 1885 1884 1883 1882 1881 1880 1879 1878 1877 1876 1875 1874 1873 1872 1871 1870 1869 1868 1867 1866 1865 1864 1863 1862 1861 1860 1859 1858 1857 1856 1855 1854 1853 1852 1851 1850 1849 1848 1847 1846 1845 1844 1843 1842 1841 1840 1839 1838 1837 1836 1835 1834 1833 1832 1831 1830 1829 1828 1827 1826 1825 1824 1823 1822 1821 1820 1819 1818 1817 1816 1815 1814 1813 1812 1811 1810 1809 1808 1807 1806 1805 1804 1803 1802 1801 1800 1799 1798 1797 1796 1795 1794 1793 1792 1791 1790 1789 1788 1787 1786 1785 1784 1783 1782 1781 1780 1779 1778 1777 1776 1775 1774 1773 1772 1771 1770 1769 1768 1767 1766 1765 1764 1763 1762 1761 1760 1759 1758 1757 1756 1755 1754 1753 1752 1751 1750 1749 1748 1747 1746 1745 1744 1743 1742 1741 1740 1739 1738 1737 1736 1735 1734 1733 1732 1731 1730 1729 1728 1727 1726 1725 1724 1723 1722 1721 1720 1719 1718 1717 1716 1715 1714 1713 1712 1711 1710 1709 1708 1707 1706 1705 1704 1703 1702 1701 1700 1699 1698 1697 1696 1695 1694 1693 1692 1691 1690 1689 1688 1687 1686 1685 1684 1683 1682 1681 1680 1679 1678 1677 1676 1675 1674 1673 1672 1671 1670 1669 1668 1667 1666 1665 1664 1663 1662 1661 1660 1659 1658 1657 1656 1655 1654 1653 1652 1651 1650 1649 1648 1647 1646 1645 1644 1643 1642 1641 1640 1639 1638 1637 1636 1635 1634 1633 1632 1631 1630 1629 1628 1627 1626 1625 1624 1623 1622 1621 1620 1619 1618 1617 1616 1615 1614 1613 1612 1611 1610 1609 1608 1607 1606 1605 1604 1603 1602 1601 1600 1599 1598 1597 1596 1595 1594 1593 1592 1591 1590 1589 1588 1587 1586 1585 1584 1583 1582 1581 1580 1579 1578 1577 1576 1575 1574 1573 1572 1571 1570 1569 1568 1567 1566 1565 1564 1563 1562 1561 1560 1559 1558 1557 1556 1555 1554 1553 1552 1551 1550 1549 1548 1547 1546 1545 1544 1543 1542 1541 1540 1539 1538 1537 1536 1535 1534 1533 1532 1531 1530 1529 1528 1527 1526 1525 1524 1523 1522 1521 1520 1519 1518 1517 1516 1515 1514 1513 1512 1511 1510 1509 1508 1507 1506 1505 1504 1503 1502 1501 1500 1499 1498 1497 1496 1495 1494 1493 1492 1491 1490 1489 1488 1487 1486 1485 1484 1483 1482 1481 1480 1479 1478 1477 1476 1475 1474 1473 1472 1471 1470 1469 1468 1467 1466 1465 1464 1463 1462 1461 1460 1459 1458 1457 1456 1455 1454 1453 1452 1451 1450 1449 1448 1447 1446 1445 1444 1443 1442 1441 1440 1439 1438 1437 1436 1435 1434 1433 1432 1431 1430 1429 1428 1427 1426 1425 1424 1423 1422 1421 1420 1419 1418 1417 1416 1415 1414 1413 1412 1411 1410 1409 1408 1407 1406 1405 1404 1403 1402 1401 1400 1399 1398 1397 1396 1395 1394 1393 1392 1391 1390 1389 1388 1387 1386 1385 1384 1383 1382 1381 1380 1379 1378 1377 1376 1375 1374 1373 1372 1371 1370 1369 1368 1367 1366 1365 1364 1363 1362 1361 1360 1359 1358 1357 1356 1355 1354 1353 1352 1351 1350 1349 1348 1347 1346 1345 1344 1343 1342 1341 1340 1339 1338 1337 1336 1335 1334 1333 1332 1331 1330 1329 1328 1327 1326 1325 1324 1323 1322 1321 1320 1319 1318 1317 1316 1315 1314 1313 1312 1311 1310 1309 1308 1307 1306 1305 1304 1303 1302 1301 1300 1299 1298 1297 1296 1295 1294 1293 1292 1291 1290 1289 1288 1287 1286 1285 1284 1283 1282 1281 1280 1279 1278 1277 1276 1275 1274 1273 1272 1271 1270 1269 1268 1267 1266 1265 1264 1263 1262 1261 1260 1259 1258 1257 1256 1255 1254 1253 1252 1251 1250 1249 1248 1247 1246 1245 1244 1243 1242 1241 1240 1239 1238 1237 1236 1235 1234 1233 1232 1231 1230 1229 1228 1227 1226 1225 1224 1223 1222 1221 1220 1219 1218 1217 1216 1215 1214 1213 1212 1211 1210 1209 1208 1207 1206 1205 1204 1203 1202 1201 1200 1199 1198 1197 1196 1195 1194 1193 1192 1191 1190 1189 1188 1187

109 / 4 US-08-836-

3	777	99.5	191	45	US-08-290-	Se
---	-----	------	-----	----	------------	----

4 111 55.5 TST 4J 03-00-067-32

CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. BOX 4433
CC CITY: HOUSTON
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word 6.0 / ASCII text output
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/836,075A
CC FILING DATE: 21 Apr 1997
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/EP95/04155
CC FILING DATE: 23 Oct 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 94870166.9
CC FILING DATE: 21 Oct 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 95870076.7
CC FILING DATE: 28 Jun 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KAMMERER, PATRICIA A.
CC REGISTRATION NUMBER: 29,775
CC REFERENCE/DOCKET NUMBER: INNS:004
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 109 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 109 AA; 12273 MW; 64913 CN;

Query Match 100.0%; Score 781; DB 74; Length 109;
Best Local Similarity 100.0%; Pred. No. 5.66e-60;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MSTNPKQKTKRNTNRRPXXXXPCCGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRRPXXXXPCCGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||

Db 61 RRQPIKAXRXEGRSWAQPGYPMPLYNCGCGMAXWLLSPRGRNMGWP 109
|||||
Qy 61 RRQPIKAXRXEGRSWAQPGYPMPLYNCGCGMAXWLLSPRGRNMGWP 109
|||||

RESULT 2
ID US-08-836-075A-28 STANDARD; PRT; 117 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 28, Application US/08836075A.
XX
CC Sequence 28, Application US/08836075A
CC GENERAL INFORMATION:
CC APPLICANT: MAERTENS, GEERT
CC APPLICANT: STUVER, LIEVEN
CC TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
CC TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DI
AGNOSTIC

CC TITLE OF INVENTION: AGENTS
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. BOX 4433
CC CITY: HOUSTON
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word 6.0 / ASCII text output
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/836,075A
CC FILING DATE: 21 Apr 1997
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/EP95/04155
CC FILING DATE: 23 Oct 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 94870166.9
CC FILING DATE: 21 Oct 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 95870076.7
CC FILING DATE: 28 Jun 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KAMMERER, PATRICIA A.
CC REGISTRATION NUMBER: 29,775
CC REFERENCE/DOCKET NUMBER: INNS:004
CC INFORMATION FOR SEQ ID NO: 28:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 117 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 117 AA; 13380 MW; 67911 CN;

Query Match 99.5%; Score 777; DB 74; Length 117;
Best Local Similarity 91.7%; Pred. No. 1.32e-59;
Matches 100; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 1 MSTNPKQKTKRNTNRRPMDVKFPGGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRRPXXXXPCCGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||

Db 61 RRQPIKAXRXEGRSWAQPGYPMPLYNCGCGMAXWLLSPRGRNMGWP 109
|||||
Qy 61 RRQPIKAXRXEGRSWAQPGYPMPLYNCGCGMAXWLLSPRGRNMGWP 109
|||||

RESULT 3
ID US-08-290-665A-195 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 195, Application US/08290665A.
XX
CC Sequence 195, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BUXH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.

CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 195:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: Z6
CC SEQUENCE 191 AA; 20774 MW; 184617 CN;

Query Match 99.5%; Score 777; DB 45; Length 191;
Best Local Similarity 90.8%; Pred. No. 1.32e-59;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKPQRTKNTNRPMVKPGCGQIVGGVYLLPRGRLGVRAVTRKTSERSQPRG 60
|||||
Qy 1 MSTNPKPQRTKNTNRPMVKPGCGQIVGGVYLLPRGRLGVRAVTRKTSERSQPRG 60
|||||
Db 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGRSPWGP 109
|||||
Qy 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGRSPWGP 109
|||||

RESULT 4
ID US-08-290-665A-192 STANDARD; PRT; 191 AA.

XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 192, Application US/08290665A.
XX
CC Sequence 192, Application US/08290665A

CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 192:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: Z8
CC SEQUENCE 191 AA; 20744 MW; 181593 CN;

Query Match 99.5%; Score 777; DB 45; Length 191;
Best Local Similarity 90.8%; Pred. No. 1.32e-59;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKPQRTKNTNRPMVKPGCGQIVGGVYLLPRGRLGVRAVTRKTSERSQPRG 60
|||||
Qy 1 MSTNPKPQRTKNTNRPMVKPGCGQIVGGVYLLPRGRLGVRAVTRKTSERSQPRG 60
|||||
Db 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGRSPWGP 109
|||||
Qy 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGRSPWGP 109
|||||

RESULT 5
ID US-07-786-169A-6 STANDARD; PRT; 120 AA.

XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
CC

DE Sequence 6, Application US/07786169A.

XX Sequence 6, Application US/07786169A

CC GENERAL INFORMATION:

CC APPLICANT: Krupka, Udo

CC APPLICANT: Stuber, Werner

CC APPLICANT: Gerken, Manfred

CC APPLICANT: Brust, Stefan

CC TITLE OF INVENTION: HCV-Specific Peptides, Agents Therefor

CC TITLE OF INVENTION: and the Use Thereof

CC NUMBER OF SEQUENCES: 44

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

CC ADDRESSEE: Dunner

CC STREET: 1300 I Street, N.W.

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: USA

CC ZIP: 20005-3315

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/07786,169A

CC FILING DATE: 19911031

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Kirschner, Michael K.

CC REGISTRATION NUMBER: 34,851

CC REFERENCE/DOCKET NUMBER: 02481-1118-00000

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 202-408-4000

CC TELEFAX: 202-408-4400

CC INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 120 amino acids

CC TYPE: AMINO ACID

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

SQ SEQUENCE 120 AA; 13619 MW; 70513 CN;

Query Match 99.2%; Score 775; DB 11; Length 120;

Best Local Similarity 89.9%; Pred. No. 2.01e-59;

Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKPQKTKNTNRPPQDVKFPGGQIVGGVYLLPRGPRGLGVTRTKTSERSQPRG 60

QY 1 MSTNPKPQKTKNTNRPPQDVKFPGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60

Db 61 RRQPIPKARPEGRTPAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPWSGP 109

QY 61 RRQPIPKAXEXGRSWAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPNWGP 109

RESULT 6

ID US-08-290-665A-158 STANDARD; PRT; 191 AA.

XX

AC xxxxxx

XX 01-JAN-1900

XX

DE Sequence 158, Application US/08290665A.

XX Sequence 158, Application US/08290665A

CC GENERAL INFORMATION:

CC APPLICANT: BUKH, J., MILLER, R.H. AND

CC APPLICANT: PURCELL, R.H.

CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

CC NUMBER OF SEQUENCES: 263

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: MORGAN & FINNEGAN

CC STREET: 345 PARK AVENUE

CC CITY: NEW YORK

CC STATE: NEW YORK

CC COUNTRY: USA

CC ZIP: 10154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: FLOPPY DISK

CC COMPUTER: IBM PC COMPATIBLE

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: WORDPERFECT 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/290,665A

CC FILING DATE: 15-AUG-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: RICHARD W. BORK

CC REGISTRATION NUMBER: 36,459

CC REFERENCE/DOCKET NUMBER: 2026-4116

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212) 758-4800

CC TELEFAX: (212) 751-6849

CC TELEX: 421792

CC INFORMATION FOR SEQ ID NO: 158:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 191 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: unknown

CC TOPOLOGY: unknown

CC ORIGINAL SOURCE:

CC ORGANISM: homosapiens

CC INDIVIDUAL ISOLATE: SW1

SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 45; Length 191;

Best Local Similarity 89.9%; Pred. No. 2.01e-59;

Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKPQKTKNTNRPPQDVKFPGGQIVGGVYLLPRGPRGLGVTRTKTSERSQPRG 60

QY 1 MSTNPKPQKTKNTNRPPQDVKFPGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60

Db 61 RRQPIPKARPEGRTPAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPWSGP 109

QY 61 RRQPIPKAXEXGRSWAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPNWGP 109

RESULT 7

ID US-08-290-665A-196 STANDARD; PRT; 191 AA.

XX

AC xxxxxx

XX

* Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPQDKFPGGQIVGGVYLLPRGPRGLGVRATRKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPQDKFPGGQIVGGVYLLPRGPRGLGVRATRKTSERSQPRG 60
|||||

Db 61 RRQIPKARRPEGRRTWAQPGYPMPLYGNECGMGWLLSPRGSRPWSGP 109
|||||
Qy 61 RRQIPKAXXEGRSWAQPGYPMPLYGNECGMGWLLSPRGSRPWSGP 109
|||||

RESULT 11
ID US-08-290-665A-157 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 157, Application US/08290665A.
XX
CC Sequence 157, Application US/08290665A
CC
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 157:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: S14

SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 45; Length 191;
Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPQDKFPGGQIVGGVYLLPRGPRGLGVRATRKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPQDKFPGGQIVGGVYLLPRGPRGLGVRATRKTSERSQPRG 60
|||||

Db 61 RRQIPKARRPEGRRTWAQPGYPMPLYGNECGMGWLLSPRGSRPWSGP 109
|||||
Qy 61 RRQIPKAXXEGRSWAQPGYPMPLYGNECGMGWLLSPRGSRPWSGP 109
|||||

RESULT 12
ID US-08-290-665A-156 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 156, Application US/08290665A.
XX
CC Sequence 156, Application US/08290665A
CC
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 156:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown

CC ORIGINAL SOURCE:
CC ORGANISM: homoeapiens
CC INDIVIDUAL ISOLATE: US11
SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 45; Length 191;
Best Local Similarity 89.9%; Pred. No. 2.0le-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||

Db 61 RRQPIKARREPGRTWAQPGYPWPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RRQPIKARREPGRTWAQPGYPWPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||

RESULT 13
ID US-08-224-973A-23 STANDARD; PRT; 200 AA.
XX
AC xxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 23, Application US/08224973A.
XX
CC Sequence 23, Application US/08224973A
CC GENERAL INFORMATION:
CC APPLICANT: Berzofsky, Jay A.
CC APPLICANT: Feinstein, Stephen
CC APPLICANT: Shirai, Mutsunori
CC TITLE OF INVENTION: Hepatitis C Virus Core Peptide for
CC TITLE OF INVENTION: Stimulation of Cytotoxic T Lymphocytes and Diagno
sis of
CC TITLE OF INVENTION: HCV Exposure
CC NUMBER OF SEQUENCES: 26
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
CC STREET: P.O. Box 747
CC CITY: Falls Church
CC STATE: Virginia
CC COUNTRY: USA
CC ZIP: 22040-0747
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/224,973A
CC FILING DATE: 08-APR-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Svensson, Leonard R.
CC REGISTRATION NUMBER: 30330
CC REFERENCE/DOCKET NUMBER: 1173-456P
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-205-8000
CC TELEFAX: 703-205-8050
CC INFORMATION FOR SEQ ID NO: 23:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 200 amino acids
CC TYPE: amino acid

CC STRANDEDNESS: not relevant
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Hepatitis C Virus
CC STRAIN: NYBC
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1..200
CC OTHER INFORMATION: /note= "HCV core protein, NYBC
CC OTHER INFORMATION: isolate, see Fig. 5"
SQ SEQUENCE 200 AA; 21771 MW; 201865 CN;

Query Match 99.2%; Score 775; DB 38; Length 200;
Best Local Similarity 89.9%; Pred. No. 2.0le-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPQDVKFGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||

Db 61 RRQPIKARREPGRTWAQPGYPWPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RRQPIKARREPGRTWAQPGYPWPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||

RESULT 14
ID US-08-710-637-1 STANDARD; PRT; 3011 AA.
XX
AC xxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 1, Application US/08710637.
XX
CC Sequence 1, Application US/08710637
CC GENERAL INFORMATION:
CC APPLICANT: CASEY, JAMES M.
CC APPLICANT: BODE, SUZANNE L.
CC APPLICANT: ZECK, BILLY J.
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: FRAIL, DONALD E.
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC TITLE OF INVENTION: PROTEINS
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/710,637
CC FILING DATE:

Feb 23 11:02

/home/fuller/feb1098/US-08-836-075A-2.rap

17

CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/144,099
CC FILING DATE:
CC APPLICATION NUMBER: US 07/830,024
CC FILING DATE: 01-JAN-1992
CC FILING DATE: 01-JAN-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5131.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-937-9556
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3011 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 3011 AA; 327108 MW; 47577590 CN;

Query Match 99.2%; Score 775; DB 58; Length 3011;
Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQRTKNTNRRPDVKFPGGQIVGGVYLLPRGPRGLGVRATRTKTSERSQPRG 60
QY 1 MSTNPKQRTKNTNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRTKTSERSQPRG 60
Db 61 RRQPIPKARPEGRWTWQPGYWPPLYGNECGGAGWLLSPRGSRPWSGP 109
QY 61 RRQPIPKAXRGEGRSWAQPGYWPPLYGNECGGAGWLLSPRGSRPNWGP 109

RESULT 15
ID US-08-710-637-2 STANDARD; PRT; 3011 AA.
XX
AC xxxxxx
XX
DT
XX 01-JAN-1900
DE Sequence 2, Application US/08710637.
CC Sequence 2, Application US/08710637
CC GENERAL INFORMATION:
CC APPLICANT: CASEY, JAMES M.
CC APPLICANT: BODE, SUZANNE L.
CC APPLICANT: ZECK, BILLY J.
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: FRAIL, DONALD E.
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC TITLE OF INVENTION: PROTEINS
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP60
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:

Feb 23 11:02

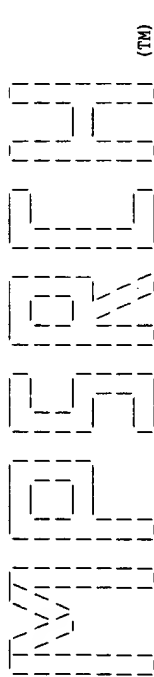
/home/fuller/feb1098/US-08-836-075A-2.rap

18

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/710,637
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/144,099
CC FILING DATE:
CC APPLICATION NUMBER: US 07/830,024
CC FILING DATE: 01-JAN-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5131.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-937-9556
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3011 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 3011 AA; 327194 MW; 47602524 CN;

Query Match 99.2%; Score 775; DB 68; Length 3011;
Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQRTKNTNRRPDVKFPGGQIVGGVYLLPRGPRGLGVRATRTKTSERSQPRG 60
QY 1 MSTNPKQRTKNTNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRTKTSERSQPRG 60
Db 61 RRQPIPKARPEGRWTWQPGYWPPLYGNECGGAGWLLSPRGSRPWSGP 109
QY 61 RRQPIPKAXRGEGRSWAQPGYWPPLYGNECGGAGWLLSPRGSRPNWGP 109

Search completed: Mon Feb 23 11:05:31 1998
Job time : 142 secs.

[illegible]

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Feb 23 11:02:22 1998; MasPar time 4.75 Seconds
486.554 Million cell updates/sec

Tabular output not generated.

Title: >US-08-836--075A-2

Description: (1-109) from US08836075A.pap

Perfect Score:

Sequence: 1 MSTNPKPQKTKRNTNRRPX.....CCGWAXWLLSPGSRPNWGP 109

Scoring table: PAM 150

Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot34

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 39.606; Variance 82.455; scale 0.480

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description	Pred. No.
		Match	Length	DB			
1	775	99.2	3011	7	POLG_HCVH	GENOME POLYPROTEIN	C 4.20e-127
2	765	98.0	3011	7	POLG_HCVI	GENOME POLYPROTEIN	C 4.05e-125
3	754	96.5	3010	7	POLG_HCVJA	GENOME POLYPROTEIN	C 6.16e-123
4	753	96.4	513	7	POLG_HCVJ2	GENOME POLYPROTEIN	C 9.72e-123
5	749	95.9	737	7	POLG_HCVJ7	GENOME POLYPROTEIN	C 6.03e-122
6	749	95.9	3033	7	POLG_HCVJB	GENOME POLYPROTEIN	C 6.03e-122
7	748	95.8	3010	7	POLG_HCVBK	GENOME POLYPROTEIN	C 9.52e-122
8	745	95.4	520	7	POLG_HCVHA	GENOME POLYPROTEIN	C 3.74e-121
9	745	95.4	520	7	POLG_HCVHK	GENOME POLYPROTEIN	C 3.74e-121
10	728	93.2	737	7	POLG_HCVJ5	GENOME POLYPROTEIN	C 8.70e-118
11	728	93.2	3010	7	POLG_HCVJT	GENOME POLYPROTEIN	C 8.70e-118
12	728	93.2	3033	7	POLG_HCVJB	GENOME POLYPROTEIN	C 8.70e-118
13	725	92.8	3010	7	POLG_HCVTA	GENOME POLYPROTEIN	C 3.41e-117

14	104	13.3	2205	7	POLN_RUBVT	NONSTRUCTURAL POLYPRO	1.10e-02
15	102	13.1	975	2	CDP_CANFA	CCAAT DISPLACEMENT PR	2.95e-02
16	98	12.5	2424	2	CB12_RABIT	BRAIN CALCIUM CHANNEL	6.95e-02
17	95	12.2	245	7	OMCM_BOVIN	ONCOSTATIN M PRECURSOR	1.11e-01
18	95	12.2	1523	9	SON3_HUMAN	SON PROTEIN (SON3)	1.71e-01
19	94	12.0	250	4	GCH1_HUMAN	GTP CYCLOHYDROLASE I	2.30e-01
20	94	12.0	420	6	LYOX_CHICK	PROTEIN-LYSINE 6-OXID	2.30e-01
21	94	12.0	494	10	VE2_HP12	REGULATORY PROTEIN E2	2.30e-01
22	94	12.0	2183	8	RPL_MEASE	RNA POLYMERASE BETA S	2.30e-01
23	94	12.0	2183	8	RPL_MEASA	RNA POLYMERASE BETA S	2.30e-01
24	93	11.9	1184	3	CYA5_CANFA	ADENYLATE CYCLASE, TY	3.08e-01
25	92	11.8	417	6	LYOX_HUMAN	PROTEIN-LYSINE 6-OXID	4.12e-01
26	92	11.8	497	10	VE2_HP20	REGULATORY PROTEIN E2	4.12e-01
27	92	11.8	662	9	SWA1_DROME	SUPPRESSOR OF WHITE A	4.12e-01
28	91	11.7	92	4	HSP1_DASRO	SERM PROTAMINE P1.	5.51e-01
29	91	11.7	241	4	GCH1_MOUSE	GTP CYCLOHYDROLASE I	5.51e-01
30	91	11.7	614	9	RUI17_HUMAN	U1 SMALL NUCLEAR RIBO	5.51e-01
31	90	11.5	888	5	KUTK_MOUSE	LEUKOCYTE TYROSINE KI	7.35e-01
32	89	11.4	60	5	HSP1_DASVI	SERM PROTAMINE P1.	9.79e-01
33	89	11.4	61	5	HSP1_SARHA	SERM PROTAMINE P1.	9.79e-01
34	89	11.4	61	5	HSP1_TRIVU	SERM PROTAMINE P1.	9.79e-01
35	89	11.4	61	5	HSP1_ANTSW	SERM PROTAMINE P1.	9.79e-01
36	89	11.4	241	4	GCH1_RAT	GTP CYCLOHYDROLASE I	9.79e-01
37	89	11.4	438	5	IE63_EBV	TRANSCRIPTIONAL REGUL	9.79e-01
38	89	11.4	576	2	CATA_RHOCA	PEROXIDASE / CATALASE	9.79e-01
39	88	11.3	216	10	VIF_HP2D2	VIROIN INFECTIVITY FA	1.30e+00
40	88	11.3	483	10	VE2_HP14	REGULATORY PROTEIN E2	1.30e+00
41	88	11.3	1618	6	NEST_HUMAN	NESTIN.	1.30e+00
42	87	11.1	101	4	H16 ENSMI	HISTONE H1-LIKE PROTE	1.72e+00
43	87	11.1	221	9	SC35_HUMAN	SPLICING FACTOR SC35	1.72e+00
44	87	11.1	512	5	IE63_HP5V1	TRANSCRIPTIONAL REGUL	2.28e+00
45	86	11.0	514	10	VE2_HP5B	REGULATORY PROTEIN E2	2.28e+00

ALIGNMENTS

RESULT	1	
ID	POLG HCVH	STANDARD; PR7; 3011 AA.
AC	P27958;	
DT	01-AUG-1992	(REL. 23, CREATED)
DT	01-AUG-1992	(REL. 23, LAST SEQUENCE UPDATE)
DT	01-FEB-1994	(REL. 26, LAST ANNOTATION UPDATE)
DE	GENOME POLYPROTEIN	(CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE	PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL	
DE	PROTEINS NS1, NS2, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA	
DE	POLYMERASE (EC 2.7.7.48) (NS5)).	
OS	HEPATITIS C VIRUS (ISOLATE H) (HCV).	
OC	VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;	
OC	HCV GROUP VIRUSES.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 92052256.	
RA	INCHAUSPE G., ZEBEDEE S., LEE D.H.H., SUGITANI M., NASOFF M.,	
RA	PRINCE A.M.;	
RL	PROC. NATIONL. ACAD. SCI. U.S.A. 88:10292-10296(1991).	
CC	-1-. FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE	
CC	HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.	
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.	
CC	-1-. SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A	
CC	LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS OF	
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF	
CC	PROTEIN C AND RNA.	
DR	EMBL; M67463; G329738; --.	
DR	PIR; A36814; GNWVCH.	

FT	CARBOHYD	233	233	POTENTIAL.
FT	CARBOHYD	234	234	POTENTIAL.
FT	CARBOHYD	250	250	POTENTIAL.
FT	CARBOHYD	305	305	POTENTIAL.
FT	CARBOHYD	417	417	POTENTIAL.
FT	CARBOHYD	423	423	POTENTIAL.
FT	CARBOHYD	430	430	POTENTIAL.
FT	CARBOHYD	448	448	POTENTIAL.
FT	NON_TER	513	513	
SQ	SEQUENCE	513 AA;	55704 MW;	6111D156 CRC32;

Query Match 96.4%; Score 753; DB 7; Length 513;
Best Local Similarity 89.0%; Pred. No. 9.72e-123;
Matches 97: Conservative 2; Mismatches 10; Indels 0; Gaps 0;

```
Db      1 mstnpkprtktnrrpqdvkfpgggqivgvyllprgprlgvratrktsersprg 60
        |||||
Ov      1 MSTNPKPORKTKTNTRRP XXXXPAGGOIVGVYLLPRGPRXGVRAATRKTSERSOPRG 60
```

61 rrqpikrrpgragwagqypwlygneaigawgwlspgsgpswgp 109
||||||| | |||:||||||| ||| |||||:||||
61 RRQPIKRRPGRAGWAGQYPWLYGNEAIGAWGWLSPGSGPSWGP 109

RESULT	5	STANDARD;	PRT;	737 AA.
ID	POLG HCVJ7			
AC	P27961;			
DT	01-AUG-1992 (REL. 23, CREATED)			
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)			
DT	01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)			
DE	GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX			
DE	PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL			
DE	PROTEINS NS1 AND NS2) (FRAGMENT).			
OS	HEPATITIS C VIRUS (ISOLATE HC-37) (HCV).			
OC	VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;			
OC	HEP. C VIRUS			

[1]
SEQUENCE FROM N.A.
RP MEDLINE: 92230232.
RX OKAMOTO H., KURAI K., OKADA S.I., YAMAMOTO K., LIZUKA H.,
RA TANAKA T., FUKUDA S., TSUDA F., MISHIRO S.;
RL VIROLOGY 188:331-341(1992).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -II- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.

[illegible]

FT	CARBOHYD	305	305	POTENTIAL.
FT	CARBOHYD	417	417	POTENTIAL.
FT	CARBOHYD	423	423	POTENTIAL.
FT	CARBOHYD	430	430	POTENTIAL.
FT	CARBOHYD	448	448	POTENTIAL.
FT	CARBOHYD	477	477	POTENTIAL.
FT	CARBOHYD	534	534	POTENTIAL.
FT	CARBOHYD	542	542	POTENTIAL.
FT	CARBOHYD	558	558	POTENTIAL.
FT	CARBOHYD	578	578	POTENTIAL.
FT	CARBOHYD	627	627	POTENTIAL.
FT	CARBOHYD	649	649	POTENTIAL.
FT	NON TER	737	737	
SQ	SEQUENCE	737 AA;	81691 MW;	86CAC08B CRC32;

Query Match	95.94;	Score 749;	DB 7;	Length 737;
Best Local Similarity	86.24;	Pred. No. 6.03e-122;		
Matches	94;	Conservative 2;	Mismatches 13;	Indels 0;
				Gaps 0;

Db	1 mstnbpqqrktkntnrnzpqavkfpgggqivggyllprgrglvratrkteerspqrg 60
Qv	1 MSTNKPQRKTKNTNRNPXXXXPGGGQIVGGVYLLPRGRPRXGVATRTKTTSRSPRG 60

```

61 rrqpkdrstgkswgkpgypwplvgnecgwagwllsprgsrptwap 109
      ||||| | :||: ||||| ||||| ||||| ||||| |||||
Ou 61 ppoobtkayvfcrcswaobcywdpivneccgagaywllsprcsrpnwcp 109

```

RESULT	6	
ID	POLG_HCVJ8	STANDARD; PRT; 3033 AA.
AC	P26661;	
DT	01-AUG-1992 (REL. 23, CREATED)	
DT	01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)	
DT	01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)	
DE	GENOME POLYPEPTIDE (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX	
DE	PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL	
DE	PROTEINS NS1, NS2, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA	
DE	DE POLYMERASE (EC 2.7.7.48) (NS5)).	
OS	HEPATITIS C VIRUS (ISOLATE HC-J8) (HCV).	
OC	VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;	

RN	[1]	SEQUENCE FROM N. A.	
RP		MEDLINE; 92230232.	
RX		OKAMOTO H., KURAI K., OKADA S.-I., YAMAMOTO K., LIZUKA H., TANAKA T.,	
RA		FUKUDA S., TSUDA F., MISHIRO S.;	
RA			
RL		VIROLOGY 188:331-341 (1992).	
CC	-!	FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE	
CC		HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.	
CC		NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.	
CC	-!	SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A	
CC		LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:	
CC		PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF	
CC		PROTEIN C AND MRNA.	
DR	EMBL;	D10988; G221609; -.	
DR	PIR;	M40250; GNMVJ8.	
DR		POLYPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN	
KW	COAT	PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE.	
KW	NONSTRUCTURAL	PROTEIN.	
FT	INIT_MET	1 1	REMOVED FROM CAPSID PROTEIN C BY THE
FT			CELLULAR AMINOPEPTIDASE.
FT	CHAIN	2 115	CAPSID PROTEIN C (POTENTIAL).
FT	CHAIN	116 191	MATRIX PROTEIN (POTENTIAL).
FT	CHAIN	192 383	MAJOR ENVELOPE PROTEIN E (POTENTIAL).

Search completed: Mon Feb 23 11:02:49 1998
Job time : 27 secs.

WATERMAN
***** (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Feb 23 21:02:30 1998; MasPar time 235.86 Seconds
981.424 Million cell updates/sec

• Tabular output not generated.

Title: >US-08-836-075A-1
Description: (1-327) from US08836075A.seq
Perfect Score: 317
N.A. Sequence: 1 ATGAGCAGCAATCTTAACCC.....CTGCGCCCAATTTGGGCGCC 327
Comp: TACTCGTCTTAGGATTGG.....GAGCGCGTTAACC CGCGG

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 1033316 seqs, 317879456 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: n-pending
1: PCT91 2: PCT92 3: PCT93 4: PCT94 5: PCT95 6: PCT96 7: PCT97
8: U75 9: U76 10: U77A 11: U77B 12: U78A 13: U78B 14: U79A
15: U79B 16: U800 17: U801 18: U802 19: U803 20: U804 21: U805
22: U806 23: U807 24: U808 25: U809 26: U810 27: U811 28: U812
29: U813 30: U814 31: U815 32: U816 33: U817 34: U818 35: U819
36: U820 37: U821 38: U822 39: U823 40: U824 41: U825 42: U826
43: U827 44: U828 45: U829 46: U830 47: U831 48: U832 49: U833
50: U834 51: U835 52: U836 53: U837 54: U838 55: U839 56: U840
57: U841 58: U842 59: U843 60: U844 61: U845 62: U846 63: U847
64: U848 65: U849 66: U85 67: U86 68: U87A 69: U87B 70: U87C
71: U88A 72: U88B 73: U88C 74: U89 75: U6000 76: U6001
77: U6002A 78: U6002B 79: U6003A 80: U6003B 81: U6004A
82: U6004B 83: U6005 84: U6006 85: NEWP 86: NEWU6 87: NEWU8

Statistics: Mean 8.813; Variance 3.802; scale 2.318

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description	Pred. No.
No.	Score					

1	317	100.0	327	87	US-08-836-	Sequence 1, Applicatio	1.88e-275
2	305	96.2	1645	9	US-07-697-	Sequence 68, Applicati	1.07e-263
3	285	89.9	573	45	US-08-290-	Sequence 114, Applicat	3.89e-244
4	285	89.9	573	45	US-08-290-	Sequence 113, Applicat	3.89e-244
5	283	89.3	327	87	US-08-836-	Sequence 5, Applicatio	3.49e-242
6	283	89.3	573	45	US-08-290-	Sequence 110, Applicat	3.49e-242
7	282	89.0	549	38	US-08-441-	Sequence 62, Applicati	3.31e-241
8	282	89.0	549	38	US-08-221-	Sequence 62, Applicati	3.31e-241
9	282	89.0	549	60	US-08-441-	Sequence 62, Applicati	3.31e-241
10	281	88.6	384	87	US-08-942-	Sequence 18, Applicati	3.14e-240
11	281	88.6	384	66	US-08-537-	Sequence 18, Applicati	3.14e-240
12	281	88.6	573	45	US-08-290-	Sequence 117, Applicat	3.14e-240
13	231	88.6	573	45	US-08-290-	Sequence 111, Applicat	3.14e-240
14	281	88.6	573	45	US-08-290-	Sequence 118, Applicat	3.14e-240
15	281	88.6	573	45	US-08-290-	Sequence 115, Applicat	3.14e-240
16	281	88.6	686	61	US-08-452-	Sequence 25, Applicati	3.14e-240
17	281	88.6	1539	63	US-08-470-	Sequence 17, Applicati	3.14e-240
18	281	88.6	1539	10	US-07-712-	Sequence 17, Applicati	3.14e-240
19	281	88.6	1863	10	US-07-712-	Sequence 14, Applicati	3.14e-240
20	281	88.6	1863	63	US-08-470-	Sequence 14, Applicati	3.14e-240
21	281	88.6	1880	11	US-07-790-	Sequence 28, Applicati	3.14e-240
22	281	88.6	2540	13	US-07-866-	Sequence 2, Applicatio	3.14e-240
23	281	88.6	2540	41	US-08-255-	Sequence 2, Applicatio	3.14e-240
24	281	88.6	9472	31	US-08-150-	Sequence 96, Applicati	3.14e-240
25	280	88.3	2433	67	US-08-612-	Sequence 49, Applicati	2.97e-239
26	279	88.0	345	9	US-07-635-	Sequence 5, Applicatio	2.82e-238
27	279	88.0	345	54	US-08-384-	Sequence 7, Applicatio	2.82e-238
28	279	88.0	573	45	US-08-290-	Sequence 112, Applicat	2.82e-238
29	279	88.0	573	45	US-08-290-	Sequence 108, Applicat	2.82e-238
30	279	88.0	573	45	US-08-290-	Sequence 7, Applicatio	2.82e-238
31	279	88.0	1167	9	US-07-635-	Sequence 7, Applicatio	2.82e-238
32	279	88.0	1167	54	US-08-383-	Sequence 7, Applicatio	2.82e-238
33	279	88.0	1167	54	US-08-384-	Sequence 9, Applicatio	2.82e-238
34	279	88.0	1167	48	US-08-324-	Sequence 9, Applicatio	2.82e-238
35	279	88.0	1499	48	US-08-324-	Sequence 3, Applicatio	2.82e-238
36	279	88.0	1665	12	US-07-819-	Sequence 1, Applicatio	2.82e-238
37	279	88.0	6039	48	US-08-324-	Sequence 11, Applicati	2.82e-238
38	279	88.0	6039	11	US-07-769-	Sequence 11, Applicati	2.82e-238
39	279	88.0	9030	9	US-07-635-	Sequence 11, Applicati	2.82e-238
40	279	88.0	9030	25	US-08-099-	Sequence 13, Applicati	2.82e-238
41	279	88.0	9030	11	US-07-769-	Sequence 13, Applicati	2.82e-238
42	279	88.0	9030	54	US-08-384-	Sequence 13, Applicati	2.82e-238
43	279	88.0	9030	11	US-07-769-	Sequence 13, Applicati	2.82e-238
44	279	88.0	9416	9	US-07-635-	Sequence 1, Applicatio	2.82e-238
45	279	88.0	9416	54	US-08-384-	Sequence 1, Applicatio	2.82e-238

ALIGNMENTS

RESULT 1
ID US-08-836-075A-1 STANDARD; DNA; UNC; 327 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 1, Application US/08836075A.
CC Sequence 1, Application US/08836075A
CC GENERAL INFORMATION:
CC APPLICANT: MAERTENS, GEERT
CC APPLICANT: STUYVEN, LIEVEN
CC TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
CC TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIA
GNOSTIC
CC TITLE OF INVENTION: AGENTS
CC NUMBER OF SEQUENCES: 207

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. BOX 4433
CC CITY: HOUSTON
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word 6.0 / ASCII text output
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/836, 075A
CC FILING DATE: 21 Apr 1997
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/EP95/04155
CC FILING DATE: 23 Oct 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 94870166.9
CC FILING DATE: 21 Oct 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 95870076.7
CC FILING DATE: 28 Jun 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KAMMERER, PATRICIA A.
CC REGISTRATION NUMBER: 29,775
CC REFERENCE/DOCKET NUMBER: INNS:1004
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 327 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
SQ Sequence 327 BP; 56 A; 102 C; 104 G; 51 T; 14 other;

Query Match 100.0%; Score 317; DB 87; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.88e-275;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 ATGACGACGAATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
QY 1 ATGACGACGAATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
Db 61 GSGTNNNNNNCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTCGGCGGAGG 120
|||||
QY 61 GSGTNNNNNNCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTCGGCGGAGG 120
|||||
Db 121 GCGCCAGGNNGGGTGTCGGCGGACTAGGAAGACTTCCGAGCGGTCAACACCTGCTGGC 180
QY 121 GCGCCAGGNNGGGTGTCGGCGGACTAGGAAGACTTCCGAGCGGTCAACACCTGCTGGC 180
|||||
Db 181 AGCGACAGCCTATCCCAAGGCTGCGGCGGAGGAGGCTTGGGCTACAGCGCGG 240
QY 181 AGCGACAGCCTATCCCAAGGCTGCGGCGGAGGAGGCTTGGGCTACAGCGCGG 240
|||||
Db 241 TATCCTTGGCCCTCTATGCAATGAGGGCTGGCGGTGGCGGCTGGCTCTGTCCTCC 300
QY 241 TATCCTTGGCCCTCTATGCAATGAGGGCTGGCGGTGGCGGCTGGCTCTGTCCTCC 300
Db 301 CGCGGCTCTCGGCCCAATTTGGGGCCCC 327

QY 301 CGCGGCTCTCGGCCCAATTTGGGGCCCC 327
|||||
RESULT 2
ID US-07-697-326A-68 STANDARD; DNA; UNC; 1645 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 68, Application US/07697326A.
CC Sequence 68, Application US/07697326A
CC GENERAL INFORMATION:
CC APPLICANT: Cha, Tai-An
CC APPLICANT: Beall, Eileen
CC APPLICANT: Irvine, Bruce
CC APPLICANT: Kolberg, Janice
CC APPLICANT: Urdea, Michael S.
CC TITLE OF INVENTION: POLYNUCLEOTIDE PROBES USEFUL FOR
CC TITLE OF INVENTION: SCREENING FOR HEPATITIS C VIRUS
CC NUMBER OF SEQUENCES: 68
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Morrison & Foerster
CC STREET: 545 Middlefield Road, Suite 200
CC CITY: Menlo Park
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94025
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/697, 326A
CC FILING DATE: 19910508
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Reed, Dianne E.
CC REGISTRATION NUMBER: 31,292
CC REFERENCE/DOCKET NUMBER: 2300-0207.00
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-327-7250
CC TELEFAX: 415-327-2951
CC TELEX: 706141
CC INFORMATION FOR SEQ ID NO: 68:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1645 base pairs
CC TYPE: NUCLEIC ACID
CC STRANDEDNESS: single
CC TOPOLOGY: linear
SQ Sequence 1645 BP; 229 A; 373 C; 384 G; 249 T; 410 other;

Query Match 96.2%; Score 305; DB 9; Length 1645;
Best Local Similarity 80.1%; Pred. No. 1.07e-263;
Matches 262; Conservative 16; Mismatches 16; Indels 0; Gaps 0;
Db 342 A-GAGCAGCATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCACAG 401
|||||
QY 1 ATGACGACGAATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
Db 402 GAGCTYAAGTCTCCGGGCGGCGGAGTGTGCTGGTGGAGTTTACCTGTCGGCGGAGG 461
|||||
QY 61 GSGTNNNNNNCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTCGGCGGAGG 120
|||||
Db 462 GCGCCYAGRTTGGGTGTGGCGGCGGCGGAGTCTTCCGAGCGGTGCGCAACCTCGCGW 521

QY	121	GGCCCCGAGNNGGGTGTGGCGGCACCTAGGAACATTC	CGCGGTCAACCTCGTGGC	180
Db	522	AGRCGWCARCCTATCCYAAAGCGDGYCRRCCGACGGY	AGGRCCTGGCCYAGCCYGGG	581
QY	181	AGGCGACAGCCATATCCCAAGGCTGTGGGTCGAGG	CGAGGTCCTGGGCTCAGCCGGG	240
Db	582	TATCCYTGCCCCCTCTAYGGYAYCAGGCGYKHS	GGRTGGCGCGGATGGCTCTCTGTCMCCC	641
QY	241	TATCCTTGCCCCCTCTATGGCAATAGAGGCTCGGG	TGGCGGNTGGCTCTGTCGCC	300
Db	642	CGYGGYTCYCGGCCTATYTGCGGCCCY	668	
QY	301	CGCGGCTCTCGGCGCAATTGGGGGCC	327	

RESULT	3
ID	US-08-290-665A-114 STANDARD; DNA; UNC; 573 dp.

AC	xxxxxx
DT	01-JAN-1900
DE	Sequence 114, Application US/08290665A.
CC	Sequence 114, Application US/08290665A
CC	GENERAL INFORMATION:
CC	APPLICANT: BUKH, J., MILLER, R.H. AND
CC	APPLICANT: PORCELL, R.H.
CC	TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC	TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC	TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC	TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC	TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC	NUMBER OF SEQUENCES: 263

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: .15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 114:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homospiegans
CC INDIVIDUAL ISOLATE: DKL

[illegible]

RESULT

```

ID   US-08-290-655A-113 STANDARD; DNA; UNC; 573 BP.
AC   xxxxxx
DT   01--JAN-1900
DE   Sequence 113, Application US/08290655A.
CC   Sequence 113, Application US/08290655A
CC   GENERAL INFORMATION:
CC   APPLICANT: BUKH, J., MILLER, R.H. AND
CC   APPLICANT: PORCELL, R.H.
CC   TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC   TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC   TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC   TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC   TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC   NUMBER OF SEQUENCES: 263
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: MORGAN & FINNEGAN
CC   STREET: 345 PARK AVENUE
CC   CITY: NEW YORK
CC   STATE: NEW YORK
CC   COUNTRY: USA
CC   ZIP: 10154
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: FLOPPY DISK
CC   COMPUTER: IBM PC COMPATIBLE
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: WORDPERFECT 5.1
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/290,655A
CC   FILING DATE: 15-AUG-1994
CC   CLASSIFICATION: 435
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: RICHARD W. BORK

```

CC	REGISTRATION NUMBER: 36,459	
CC	REFERENCE/DOCKET NUMBER: 2026-4116	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (212) 758-4800	
CC	TELEFAX: (212) 751-6849	
CC	TELEX: 421792	
CC	INFORMATION FOR SEQ ID NO: 113:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 573 base pairs	
CC	TYPE: nucleic acid	
CC	STRANDEDNESS: single	
CC	TOPOLOGY: linear	
CC	ORIGINAL SOURCE:	
CC	ORGANISM: homopiens	
CC	INDIVIDUAL ISOLATE: P10	
CC	Sequence 573 BP; 96 A; 185 C; 179 G; 113 T; 0 other;	
CC	Query Match 89.9%; Score 285; DB 45; Length 573;	
CC	Best Local Similarity 90.9%; Pred. No. 3.89e-244;	
CC	Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;	
db	1 ATGACGACGAATCCTAAACCTCAAGAAAAACAAAGCTAAACACCAACCGCGCCACAG 60	
cc		
cc	1 ATGACGACGAATCCTAAACCTCAAGAAAAACAAAGCTAAACACCAACCGCGCCCTCAK 60	
cc		
db	61 GAGCTCAAGTTCGCGGGCGGTGTCAGATGCTTGTGGAGCTTACCTGTCGGCGGAGG 120	
cc		
cc	61 GCGGTNNNNNNCCGGTGGCGGTGAGATCGTGTGGAGCTTACCTGTCGGCGGAGG 120	
cc		
db	121 GCGCCGACGTTGGTGTGCGCGCGACTAGGAGACCTCCGACGGCTCCCAACCTCGTGA 180	
cc		
cc	121 GCGCCGACGNNGGTGTGCGCGCGACTAGGAGACCTCCGACGGTCCAAACCTCGTGGC 180	
cc		
db	181 AGGCGACACATATCCCAAGGCTCGCGCGCCGACGAGGCGAGGGCTGGGCTAGAGCCGG 240	
cc		
cc	181 AGGCGACAGCTATCCCAAGGCTCGYCGYCGGAGGAGGCTCTGGGCTAGAGCCGGG 240	
cc		
db	241 TACCTTTGGCCCTCTATGGCAATGAGGCTTGGGTTGGCAGGATGGCTCTGTCAACC 300	
cc		
cc	241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGGTGGCGGGTGCTCTGTCTCCCC 300	
cc		
db	301 CTGGCTCTCGGCCCTAGTTGGGGCCCC 327	
cc		
cc	301 CGGGCTCTCGGCCCAATTGGGGCCCC 327	
cc		
cc	RESULT 5	
cc	US-08-836-075A-5 STANDARD; DNA; UNC; 327 BP.	
cc	xxxxx	
cc	01-JAN-1900	
cc	Sequence 5, Application US/08836075A.	
cc	Sequence 5, Application US/08836075A	
cc	GENERAL INFORMATION:	
cc	APPLICANT: MAERTENS, GEERT	
cc	APPLICANT: STUYVER, LIEVEN	
cc	TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES	
cc	TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIA	
cc	GNOSTIC	
cc	TITLE OF INVENTION: AGENTS	
cc	NUMBER OF SEQUENCES: 207	
cc	CORRESPONDENCE ADDRESS:	
cc	ADDRESSEE: ARNOLD, WHITE & DURKEE	
cc	STREET: P.O. BOX 4433	
cc	CITY: HOUSTON	

CC	STATE:	TEXAS
CC	COUNTRY:	USA
CC	ZIP:	77210-4433
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	Microsoft Word 6.0 / ASCII text output
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/836,075A
CC	FILING DATE:	21 Apr 1997
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	PCT/EP95/04155
CC	FILING DATE:	23 Oct 1995
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	EP 94870166.9
CC	FILING DATE:	21 Oct 1994
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	EP 95870076.7
CC	FILING DATE:	28 Jun 1995
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	KAMMERER, PATRICIA A.
CC	REGISTRATION NUMBER:	29,775
CC	REFERENCE/DOCKET NUMBER:	INNS:004
CC	INFORMATION FOR SEQ ID NO:	5:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	327 base pairs
CC	TYPE:	nucleic acid
CC	STRANDEDNESS:	single
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	cDNA
CC	HYPOTHETICAL:	NO
CC	ANTI--SENSE:	NO
SQ	Sequence	327 bp; 62 A; 105 G; 103 C; 54 T; 3 other;
Query Match 89.3%; Score 283; DB 87; Length 327;		
Best Local Similarity 90.5%; Pred.No. 3.49e-242;		
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;		
Dn	1	ATGACGACGAATCCTTAACCTCAAGAAAACCAAACGTTAAACAACCGCCGCCACAG 60
Qy	1	ATGACGACGAATCCTTAAACTCAAGAAAACCAAACGTTAAACAACCGCCGCCCTCAK 60
Dn	61	GACGTCAAGNTCCGGGTGGTGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGGCGAGG 120
Qy	61	GSSTNNNNNNNCCGGGTGGGTGGTGCAGATCGTTGGTGGAGTTTACCTGTTGCCGGCGAGG 120
Dn	121	GGCCCCAGGTGGGTGTCGGCGGCACACGAGAAGACTTCCCACGGGTCCAGCCTCGTCAC 180
Qy	121	GGCCCCAGGNNGGTGTGGCGCGCATCCAGAGACTTCCGACGGGTCAACAACCTGTGGC 180
Dn	181	AGGCGCACAGCCTATTCTTAAGGCTCGCCAGTCCGATGGCAGNNCTTGGGTCCAGCAGG 240
Qy	181	AGGCGACAGCCTATCCCAAGGCTCGYCCGACGAGGAGCTCTCTGGGTTCAGCCCGGG 240
Dn	241	CATCCCTGGGCCCTCTATGGCAATGAGGCTCGCGATGGCGGGATGGCTCTCTGTCCGCC 300
Qy	241	TATCCTTGGCCCTCTATGCGCAATGAGGCTCGCGGTGGCGGNGTGGCTCTGTCCGCC 300
Dn	301	CGCGGCTCTGGGCCAGATTGGGGCCCC 327
Qy	301	CGCGGCTCTGGGCCAAATTGGGGCCCC 327

```
RESULT 6
ID US-08-290-665A-110 STANDARD; DNA; UNC; 573 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 110, Application US/08290665A.
CC Sequence 110, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290, 665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 110:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: hom sapiens
CC INDIVIDUAL ISOLATE: S45
SQ Sequence 573 BP; 95 A; 189 C; 177 G; 112 T; 0 other;

Query Match 89.3%; Score 283; DB 45; Length 573;
Best Local Similarity 90.5%; Pred. No. 3.49e-242;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCTTAAACCTCAAGACAAACCAACGTAACACGACCGCGCCACAG 60
|: |||
Qy 1 ATGACGACGAATCTTAAACCTCAAGACAAACCAACGTAACACGACCGCGCCACAG 60
|: |||
Db 61 GAGGTCAAGTTCGGGGTGGGGTGCAGATCGTTGGTGGAGTTTACCTGTTCGGCCGAGG 120
|: |||
Qy 61 GGSGTNNNNNNCCGGGTGGCGGTGCAGATCGTTGGTGGAGTTTACCTGTTCGGCCGAGG 120
|: |||
Db 121 GGCCCCAGGTGGGTGGTGGCGCGACTAGGACACTCCGAGCGGTCAACACTCGTGG 180
|: |||
```

```
Qy 121 GCGCCGACGNNGGGTGTGCGCGCGACTAGCAAGACTTCCGAGCGGTCAACACTCGTGGC 180
Db 181 CGGCGACAACTATCCCAAGGCTCGCGCGCGCGAGGCGAGGCGCTGGGCGCAGCCCGGG 240
|: |||
Qy 181 AGGCGACAGCCTATCCCAAGGCTCGCGCGCGCGAGGCGAGGCTCGGCTCAGCCCGGG 240
|: |||
Db 241 CATCTTGGCCCTCTATGCAATGAGGCTTGGGGTGGGCGAGGATGGCTCTCTGTGACCC 300
|: |||
Qy 241 TATCTTGGCCCTCTATGCAATGAGGCTTGGGGTGGGCGAGGATGGCTCTCTGTGACCC 300
|: |||
Db 301 CGTGCTCGCGCGCTAGTGGGGGCC 327
|: |||
Qy 301 CGCGCTCGCGCGCAATGGGGGCC 327
|: |||

RESULT 7
ID US-08-441-971-62 STANDARD; DNA; UNC; 549 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 62, Application US/08441971.
CC Sequence 62, Application US/08441971
CC GENERAL INFORMATION:
CC APPLICANT: Tai-An Cha
CC TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
CC TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
CC NUMBER OF SEQUENCES: 147
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
CC STREET: 600 Atlantic Avenue
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 5.25 inch
CC COMPUTER: IBM compatible
CC OPERATING SYSTEM: MS-DOS Version 3.3
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/441,971
CC FILING DATE: 16-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/221,653
CC FILING DATE:
CC APPLICATION NUMBER: US/07/881,528
CC FILING DATE:
CC APPLICATION NUMBER: 07/697,326
CC FILING DATE: 8 May 1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Janiuk, Anthony J.
CC REGISTRATION NUMBER: 29,809
CC REFERENCE/DOCKET NUMBER: C0772/7000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 720-3500
CC TELEFAX: (617) 720-2441
CC TELEX: EZEKIEL
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 549 nucleotides
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA
```


CC	ORIGINAL SOURCE:	
CC	INDIVIDUAL ISOLATE: spl	
SQ	Sequence 549 BP; 92 A; 175 C; 173 G; 109 T; 0 other;	
	Query Match	89.0%; Score 282; DB 60; Length 549;
	Best Local Similarity	90.5%; Pred. No. 3.31e-241;
	Matches	295; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
Db	1	ATGAGCAGCAATCTCTAAACCTCAAGAGAAAACCAAACTAAACCAACCGCGCCGCCACAG 60
Qy	1	ATGAGCAGCAATCTCTAAACCTCAAGAGAAAACCAAACTAAACCAACCGCGCCGCCCTCAK 60
Db	61	GAGCTCAAGTTCCTCCGGCGGTGCTCAGATCGTTGGTGGAGTTTACCTGTTCCGCGCGCAGG 120
Qy	61	GGSGTNNNNNNCCGGGTGGCGGTTCAGATCGTTGGTGGAGTTTACCTGTTCCGCGCGCAGG 120
Db	121	GGCCCAAGTTGGGTGTGGCGCGCACTAGAGAGACTCCCGAGCGGTCCGCAACCTCGTGA 180
Qy	121	GGCCCAAGNNGGTGTGGCGCGCACTAGAGAGACTCCCGAGCGGTTCGAGACCTCGTGGC 180
Db	181	AGGCGACAACCTATCCCAAGGCTCGCGCGCCGCGAGGCGCCCTGGGCTCAGCCCGGG 240
Qy	181	AGGCGAGAGCTATCCCAAGGCTCGYCGYCCGAGGGCAGGTCTTGGGCTCAGCCCGGG 240
Db	241	TATCTTGGCCCTCTATGGCAATCAGGCTCTGGGTGGCGCAGATGGCTCTGTCAACC 300
Qy	241	TATCTTGGCCCTCTATGGCAATGAGGCTCGGGGTGGCGGAGTGGCTCTGTCTCTGTCGCC 300
Db	301	CGCGGCTCTCGGCTAGCTGGGGCCC 326
Qy	301	CGCGGCTCTCGGCGCAATTGGGGCCC 326
RESULT	8	
ID	US-08-221-653-62	STANDARD; DNA; UNC; 549 BP.
AC	xxxxxx	
DE	01-JAN-1900	
DT	Sequence 62,	Application US/08221653.
CC	Sequence 62,	Application US/08221653
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Tai-An Cha
CC	TITLE OF INVENTION:	HCV GENOMIC SEQUENCES FOR
CC	TITLE OF INVENTION:	DIAGNOSTICS AND THERAPEUTICS
CC	NUMBER OF SEQUENCES:	147
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Wolf, Greenfield & Sacks, P.C.
CC	STREET:	600 Atlantic Avenue
CC	CITY:	Boston
CC	STATE:	Massachusetts
CC	COUNTRY:	USA
CC	ZIP:	02210
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Diskette, 5.25 inch
CC	COMPUTER:	IBM compatible
CC	OPERATING SYSTEM:	MS-DOS Version 3.3
CC	SOFTWARE:	WordPerfect 5.1
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/221,653
CC	FILING DATE:	
CC	CLASSIFICATION:	435
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/07/881,528
CC	FILING DATE:	
CC	APPLICATION NUMBER:	07/697,326

CC	FILING DATE: 8 May 1991	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Janiuk, Anthony J.	
CC	REGISTRATION NUMBER: 29,809	
CC	REFERENCE/DOCKET NUMBER: C0772/7000	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (617) 720-3500	
CC	TELEFAX: (617) 720-2441	
CC	TELEX: EZEKIEL	
CC	INFORMATION FOR SEQ ID NO: 62:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 549 nucleotides	
CC	TYPE: nucleic acid	
CC	STRANDEDNESS: single	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: DNA	
CC	ORIGINAL SOURCE:	
CC	INDIVIDUAL ISOLATE: spl	
CC	Sequence 549 BP; 92 A; 175 C; 173 G; 109 T; 0 other;	
CC	Query Match 89.0%; Score 282; DB 38; Length 549;	
CC	Best Local Similarity 90.5%; Pred.No. 3.31e-241;	
CC	Matches 295; Conservative 4; Mismatches 27; Indels 0; Gaps 0;	
DB	1 ATGAGCAGCATCTTAACCTCAAGAAAAACCAACGTAAACCAACCGCGCGCCACAG 60	
QY	1 ATGAGCAGCATCTTAACCTCAAGAAAAACCAACGTAAACCAACCGCGCGCCCTCAK 60	
DB	61 GAGCTCAAGTTCCCGCGCGGTGTCAGATCTGTTGTCGAGTTTACCTCTTGC CGCGCAGG 120	
QY	61 GGGGTNNNNNNCCGGGTGGGTGTCAGATCTGTTGTCGAGTTTACCTCTTGC CGCGCAGG 120	
DB	121 GGCCCGACGTTGGGTGTCGCGGACCTAGGAAGACTTCCGACGCTCGCAACTCTGTGA 180	
QY	121 GGCCCGACGNNNGTCTCGCGCGACCTAGCAAGACTTCCGACGCGTCAACCTGTGGC 180	
DB	181 AGCGGCAACCTATCCCAAGGCTCGCCGCGCCCGAGGAGGGCTGGGCTAGCCCGGG 240	
QY	181 AGCGGACAGCTATCCCAAGGCTCGYCGGTCGAGGCGGAGGTCCTGGGCTCAGCCCGG 240	
DB	241 TATCCTTGGCCCTCTATGCAATGAGGGTCTGGGGTGGGAGAGTGGCTCTGTTCACC 300	
QY	241 TATCCTTGGCCCTCTATGCAATGAGGGCTGGGGTGGCGGGTGGCTCTGTCTCCCC 300	
DB	301 CGCGGCTCTCGGCGTACTGGGGGCC 326	
QY	301 CGCGGCTCTCGGCGCAATTGGGGGCC 326	
RESULT	9	
ID	US-08-141-970-62 STANDARD; DNA; UNC; 549 BP.	
AC	xxxxxx	
DT	01-JAN-1900	
DE	Sequence 62, Application US/08441970.	
CC	Sequence 62, Application US/08441970	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Tai-An Cha	
CC	TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR	
CC	TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS	
CC	NUMBER OF SEQUENCES: 147	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Wolf, Greenfield & Sacks, P.C.	
CC	STREET: 600 Atlantic Avenue	
CC	CITY: Boston	

CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 5.25 inch
CC COMPUTER: IBM compatible
CC OPERATING SYSTEM: MS-DOS Version 3.3
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/441,970
CC FILING DATE: 16-MAY-1995
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/881,528
CC FILING DATE: 08-MAY-1992
CC APPLICATION NUMBER: 07/697,326
CC FILING DATE: 8 May 1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Janiak, Anthony J.
CC REGISTRATION NUMBER: 29,809
CC REFERENCE/DOCKET NUMBER: C0772/7000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 720-3500
CC TELEFAX: (617) 720-2441
CC TELEX: EZEKIEL
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 549 nucleotides
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA
CC ORIGINAL SOURCE:
CC INDIVIDUAL ISOLATE: spl
SQ Sequence 549 BP; 92 A; 175 C; 173 G; 109 T; 0 other;

Query Match 89.0%; Score 282; DB 60; Length 549;
Best Local Similarity 90.5%; Pred. No. 3.31e-241;
Matches 295; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
Db 1 ATGAGCAGCAATCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
|||||
QY 1 ATGAGCAGCAATCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTAK 60
61 GAGCTCAAGTTCCGGCGGCTGCAGATCGTTGGTGGAGTTTACCTGTCCGCCGCGAGG 120
|: || |||| || |||||
QY 61 GGGTNNNNNNCCGGGTGGCGGTGCAGATCGTTGGTGGAGTTTACCTGTCCGCCGCGAGG 120
121 GGGCCCGAGTTGGCTGTCGGCGGCTAGGACACTCCGAGCGGTCCGACACTCGTGA 180
|||||
QY 121 GGGCCCGAGTTGGCTGTCGGCGGCTAGGACACTCCGAGCGGTCCGACACTCGTGGG 180
181 AGCGCAGCACTATCCCAAGCTCGCCGCGGCGGCGGCGGCGGCTGGGCTCAGCCCGG 240
|||||
QY 181 AGCGCAGCACTATCCCAAGCTCGCGGCTCGCGGCGGCGGCGGCGGCTGGGCTCAGCCCGG 240
241 TATCTTGGCCCTCTATGGCAATGAGGGTCTGGGGTGGGCGAGATGGCTCTGTCAACC 300
|||||
QY 241 TATCTTGGCCCTCTATGGCAATGAGGGTCTGGGGTGGGCGAGATGGCTCTGTCTCCCC 300
301 CGGGCTCTCGGCGCTAGCTGGGCGCC 326
QY 301 CGGGCTCTCGGCGCAATTTGGGCGCC 326

RESULT 10
ID US-08-942-483-18 STANDARD; DNA; UNC; 384 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 18, Application US/08942483.
CC Sequence 18, Application US/08942483
CC GENERAL INFORMATION:
CC APPLICANT: Joong, Myung
CC APPLICANT: Choi, Deog Young
CC APPLICANT: Kim, Chun Hyung
CC APPLICANT: Yang, Jae Young
CC APPLICANT: Kim, In Soo
CC APPLICANT: Choi, Dong Seob
CC TITLE OF INVENTION: DIAGNOSTIC KIT AND METHOD
CC TITLE OF INVENTION: FOR THE SIMULTANEOUS DIAGNOSIS OF HEPATITIS B AND
CC NUMBER OF SEQUENCES: 35
CC TITLE OF INVENTION: C
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: PENNIE & EDMONDS LLP
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: NY
CC COUNTRY: U.S.A
CC ZIP: 10036-2711
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/942,483
CC FILING DATE: 02-OCT-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/537,814
CC FILING DATE: 24-OCT-1995
CC APPLICATION NUMBER: KR-93-7231
CC FILING DATE: 28-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jones III, Harry C.
CC REGISTRATION NUMBER: 20,280
CC REFERENCE/DOCKET NUMBER: 8512-036-999
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 18:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 384 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC FEATURE:
CC OTHER INFORMATION: KHCV CORE14, Fig. 1
SQ Sequence 384 BP; 76 A; 120 C; 121 G; 67 T; 0 other;

Query Match 88.6%; Score 281; DB 87; Length 384;
Best Local Similarity 90.5%; Pred. No. 3.14e-240;
Matches 296; Conservative 3; Mismatches 28; Indels 0; Gaps 0;
Db 1 ATGAGCAGCAATCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
|||||
QY 1 ATGAGCAGCAATCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTAK 60

CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290, 665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 117:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: IND3
SQ Sequence 573 BP; 100 A; 184 C; 174 G; 115 T; 0 other;
Query Match 88.6%; Score 281; DB 45; Length 573;
Best Local Similarity 90.2%; Pred. No. 3.14e-240;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;
Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
Db 61 GAGCTCAAGTTCGCGGGGTGCCAGATCGTTGGTGAGTTTACCTGTGCGCGCGCAGG 120
Qy 61 GCGTNNNNNNCCGGGTGGCGGTGCAGCTGTTGGTGAGTTTACCTGTGCGCGCGCAGG 120
Db 121 GCGCCAGGTTGGGTGGCGCGGACTAGGAGACTTCCGAGCGGTGCGAACCTCGTGA 180
Qy 121 GCGCCAGGNNGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTGCGAACCTCGTGGC 180
Db 181 AGCGCACACCTATCCCAAGGTCGCGCGGAGGCTAGGCGCTGGGCTCAGCCCGGG 240
Qy 181 AGCGCACAGCTATCCCAAGGTCGCGCGGAGGCTAGGCGCTGGGCTCAGCCCGGG 240
Db 241 TACCTTGGCCCTCTATGCAATGAGGGCTTGGGTGGGAGGATGGCTCTGTGCACC 300
Qy 241 TATCCTGGCCCTCTATGCAATGAGGGCTTGGGTGGGAGGATGGCTCTGTGCACC 300
Db 301 CGCGGTCTCGGCCTAGTTGGGCCCC 327
Qy 301 CGCGGTCTCGGCCTAGTTGGGCCCC 327
RESULT 13
ID US-08-290-665A-111 STANDARD; DNA; UNC; 573 BP.
AC xxxxx
DT 01-JAN-1980
DE Sequence 111, Application US/08290665A.
CC Sequence 111, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R. H. AND

CC APPLICANT: PORCELL, R. H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290, 665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 111:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: D1
SQ Sequence 573 BP; 96 A; 184 C; 177 G; 116 T; 0 other;
Query Match 88.6%; Score 281; DB 45; Length 573;
Best Local Similarity 90.2%; Pred. No. 3.14e-240;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;
Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
Db 61 GAGCTCAAGTTCGCGGGGTGGTCAGATCGTTGGTGAGTTTACCTGTGCGCGCGCAGG 120
Qy 61 GCGTNNNNNNCCGGGTGGCGGTGCAGCTGTTGGTGAGTTTACCTGTGCGCGCGCAGG 120
Db 121 GCGCCAGGTTGGGTGGCGCGGACTAGGAGACTTCCGAGCGGTGCGAACCTCGTGA 180
Qy 121 GCGCCAGGNNGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTGCGAACCTCGTGGC 180
Db 181 AGCGCACACCTATCCCAAGGTCGCGCGGAGGCTAGGCGCTGGGCTCAGCCCGGG 240
Qy 181 AGCGCACAGCTATCCCAAGGTCGCGCGGAGGCTAGGCGCTGGGCTCAGCCCGGG 240
Db 241 TACCTTGGCCCTCTATGCAATGAGGGCTTGGGTGGGAGGATGGCTCTGTGCACC 300
Qy 241 TATCCTGGCCCTCTATGCAATGAGGGCTTGGGTGGGAGGATGGCTCTGTGCACC 300

Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGGCGTGGCGGGTGGCTCTGTCTCCCC 300

Db 301 CGCGGCTCCCGGCTAGTTGGGCCCC 327

Qy 301 CGCGGCTCCCGGCAATGGGCCCC 327

RESULT 14

ID US-08-290-665A-118 STANDARD; DNA; UNC; 573 BP.

AC xxxxx

DT 01-JAN-1900

DE Sequence 118, Application US/08290665A.

CC Sequence 118, Application US/08290665A

CC GENERAL INFORMATION:

CC APPLICANT: BUKH, J., MILLER, R.H. AND

CC APPLICANT: PURCELL, R.H.

CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

CC NUMBER OF SEQUENCES: 263

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: MORGAN & FINNEGAN

CC STREET: 345 PARK AVENUE

CC CITY: NEW YORK

CC STATE: NEW YORK

CC COUNTRY: USA

CC ZIP: 10154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: FLOPPY DISK

CC COMPUTER: IBM PC COMPATIBLE

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: WORDPERFECT 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/290, 665A

CC FILING DATE: 15-AUG-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: RICHARD W. BORK

CC REGISTRATION NUMBER: 36,459

CC REFERENCE/DOCKET NUMBER: 2026-4116

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212) 758-4800

CC TELEFAX: (212) 751-6849

CC TELEX: 421792

CC INFORMATION FOR SEQ ID NO: 118:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 573 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC ORIGINAL SOURCE:

CC ORGANISM: hominids

CC INDIVIDUAL ISOLATE: IND8

SQ Sequence 573 BP; 97 A; 185 C; 177 G; 114 T; 0 other;

Query Match 88.6%; Score 281; DB 45; Length 573;

Best Local Similarity 90.2%; Pred. No. 3.14e-240;

Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCCTTAACCTCAAGAAACCAACGATACACCGCGGCCACAG 60

|||||

1 ATGACGACGAATCCTTAACCTCAAGAAACCAACGATACACCGCGGCCCTCAK 60

Db 61 GACGTCAAGTTCCCGGCGGTGGCCACAGATCGTTGGTGGAGTTTACCTTGGCGGCGCAGG 120

|||||

Qy 61 GGGTNNNNNNCCGGGTGGCGGTGACATCGTTGGTGGAGTTTACCTTGGCGGCGCAGG 120

Db 121 GGCCCAAGTTGGGTGGCGGCGACATAGGAAGACTTCGAGCGGTGCGCAACCTGCTGGA 180

|||||

Qy 121 GGCCCAAGNNGGGTGGCGGCGCACTAGGAAGACTTCGAGCGGTGCGCAACCTGCTGGA 180

Db 181 AGCGCACACCTATCCCAAGGCTCGCGGCGGAGGTAGGCGCTGGGCTCAGGCCCGG 240

|||||

Qy 181 AGCGCACACCTATCCCAAGGCTCGCGGCGGAGGTAGGCGCTGGGCTCAGGCCCGG 240

Db 241 CACCTTGGCCCTCTATGGCAATGAGGCTTGGGTTGGCGAGGATGGCTCTGTGACCC 300

|||||

Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTTGGGTTGGCGGAGGCTCTGTGACCC 300

Db 301 CGCGGCTCTCGGCTAGTTGGGCCCC 327

|||||

Qy 301 CGCGGCTCTCGGCTAGTTGGGCCCC 327

RESULT 15

ID US-08-290-665A-115 STANDARD; DNA; UNC; 573 BP.

AC xxxxx

DT 01-JAN-1900

DE Sequence 115, Application US/08290665A.

CC Sequence 115, Application US/08290665A

CC GENERAL INFORMATION:

CC APPLICANT: BUKH, J., MILLER, R.H. AND

CC APPLICANT: PURCELL, R.H.

CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

CC NUMBER OF SEQUENCES: 263

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: MORGAN & FINNEGAN

CC STREET: 345 PARK AVENUE

CC CITY: NEW YORK

CC STATE: NEW YORK

CC COUNTRY: USA

CC ZIP: 10154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: FLOPPY DISK

CC COMPUTER: IBM PC COMPATIBLE

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: WORDPERFECT 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/290, 665A

CC FILING DATE: 15-AUG-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: RICHARD W. BORK

CC REGISTRATION NUMBER: 36,459

CC REFERENCE/DOCKET NUMBER: 2026-4116

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212) 758-4800

CC TELEFAX: (212) 751-6849

CC TELEX: 421792

CC INFORMATION FOR SEQ ID NO: 115:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 573 base pairs

CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:

CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: T10

SQ Sequence 573 BP; 99 A; 181 C; 176 G; 117 T; 0 other;

Query Match 88.6%; Score 281; DB 15; Length 573;
Best Local Similarity 90.2%; Pred. No. 3.14e-240;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACACGTAACACCAACCGCGCCACAG 60
|||

Qy 1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACACGTAACACCAACCGCGCCCAK 60
|||

Db 61 GAGGTCAAGTTCCGGCGGTGTCAGATCGTTGGTGGAGTTTACCTGTCCGCGCAGG 120
|||

Qy 61 GSGTNNNNNNCCGGGTGGCGTCAGATCGTTGGTGGAGTTTACCTGTCCGCGCAGG 120
|||

Db 121 GCGCCAGGTGGGTGTCGCGCGACTAGGAGACTTCCGAGCGGTCCGAACTCGTGA 180
|||

Qy 121 GCGCCAGGNNGGTGTGCGCGCGACTAGGAGACTTCCGAGCGGTCCGAACTCGTGC 180
|||

Db 181 AGGCGACAGCTATCCCAAGGCTCCGACGCGGCGGCGCTGGGCTCAGCCCGGG 240
|||

Qy 181 AGGCGACAGCTATCCCAAGGCTCGYCGYCGGAGGCGAGGCTCTGGGCTCAGCCCGGG 240
|||

Db 241 TACCTTGGCCCTCTATGCGCAATGAGGGCATGGGCTGGGAGGATGGCTCTCTACCC 300
|||

Qy 241 TATCCTTGGCCCTCTATGCAATCAGGGCTGGGCTGGGCGGENTGGCTCTCTCCCC 300
|||

Db 301 CGTGCTCCCGGCTAGTTGGGCCCC 327
|||

Qy 301 CGCGGCTCTCGGCCCAATTGGGCCCC 327
|||

Search completed: Mon Feb 23 21:06:34 1998
Job time : 244 secs.

Best Local Similarity 90.8%; Pred. No. 1.04e-106;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkfpgggqivggvylprtgprlgvratrktseersqrg 60
|||||
Qy 1 MSTNPKPQRKTNRNTNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSEERSQPRG 60
|||||

Db 61 rrqpikarrsegwagpypwplynegcgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRXEGRSWAQPGYPWPLYNCGCGWAXWLLSPRGRPNWGP 109
|||||

RESULT 2
ENTRY S41364 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 4, N3)
CONTAINS (fragment)
ORGANISM #formal_name hepatitis C virus
#variety genotype_4, N3
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41364
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

CLASSIFICATION #molecule_type genomic RNA
#residues 1-115 #label VAN
#cross-references EMBL:229467
#experimental_source genotype 4, N3

KEYWORDS #superfamily hepatitis C virus genome polyprotein
#variety capsid protein; core protein; polyprotein

FEATURE 1-115
#product core protein #status predicted #label MAT
SUMMARY #length 115 #checksum 8983

Query Match 99.5%; Score 777; DB 9; Length 115;
Best Local Similarity 90.8%; Pred. No. 1.04e-106;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkfpgggqivggvylprtgprlgvratrktseersqrg 60
|||||
Qy 1 MSTNPKPQRKTNRNTNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSEERSQPRG 60
|||||

Db 61 rrqpikarrsegwagpypwplynegcgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRXEGRSWAQPGYPWPLYNCGCGWAXWLLSPRGRPNWGP 109
|||||

RESULT 3
ENTRY S41363 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 4, N2)
CONTAINS (fragment)
ORGANISM core protein
#variety genotype 4, N2
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41363
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

Accession S41363
#molecule_type genomic RNA
#residues 1-117 #label VAN
#cross-references EMBL:229466
#experimental_source genotype 4, N2

CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS #capsid protein; core protein; polyprotein

FEATURE 1-117
#product core protein #status predicted #label MAT
SUMMARY #length 117 #checksum 41

Query Match 99.5%; Score 777; DB 9; Length 117;
Best Local Similarity 91.7%; Pred. No. 1.04e-106;
Matches 100; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkfpgggqivggvylprtgprlgvratrktseersqrg 60
|||||
Qy 1 MSTNPKPQRKTNRNTNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSEERSQPRG 60
|||||

Db 61 rrqpikarrsegwagpypwplynegcgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRXEGRSWAQPGYPWPLYNCGCGWAXWLLSPRGRPNWGP 109
|||||

RESULT 4
ENTRY S41341 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 1, N1)
CONTAINS (fragment)
ORGANISM core protein
#formal_name hepatitis C virus
#variety genotype 1, N1
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41341
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

CLASSIFICATION #molecule_type genomic RNA
#residues 1-112 #label VAN
#cross-references EMBL:229444
#experimental_source genotype 1, N1

KEYWORDS #superfamily hepatitis C virus genome polyprotein
#variety capsid protein; core protein; polyprotein

FEATURE 1-112
#product core protein #status predicted #label MAT
SUMMARY #length 112 #checksum 9

Query Match 99.2%; Score 775; DB 9; Length 112;
Best Local Similarity 89.9%; Pred. No. 2.23e-106;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkfpgggqivggvylprtgprlgvratrktseersqrg 60
|||||
Qy 1 MSTNPKPQRKTNRNTNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSEERSQPRG 60
|||||

Db 61 rrqpikarrsegwagpypwplynegcgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRXEGRSWAQPGYPWPLYNCGCGWAXWLLSPRGRPNWGP 109
|||||

RESULT 5
ENTRY S41345 #type fragment

TITLE genome polyprotein - hepatitis C virus (genotype 1, N3)
(fragment)
CONTAINS core protein
ORGANISM #formal_name hepatitis C virus
#variety genotype 1, N3
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997
ACCESSIONS S41345
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
#accession S41345
##molecule_type genomic RNA
##residues_ 1-115 #label VAN
##cross-references EMBL:229448
##experimental_source genotype 1, N3
CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS capsid protein; core protein; polyprotein
FEATURE 1-115
#product core protein #status predicted #label MAT
SUMMARY #length 115 #checksum 9357

Query Match 99.2%; Score 775; DB 9; Length 115;
Best Local Similarity 89.9%; Pred. No. 2.23e-106;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnpkqrktrkntnrpqdvkfgggqivgvyllprgprlgvratrktsersqprg 60
|||||
Qy 1 MSTNPKQRKTRKNTNRPRXXXXPGGGIVGGVLLPRGPRXGVRATRKTSERSQPRG 60

Db 61 rrcqpkarrpgrtwagpywplyngcgwagwllsprgrpswqp 109
|||||
Qy 61 RRQPIKAXRKGSRWAQPGYPMPLYNGEGCGWALLSPRGRPNWGP 109

RESULT 6
ENTRY #molecule_type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 1, N4)
(fragment)
CONTAINS core protein
ORGANISM #formal_name hepatitis C virus
#variety genotype 1, N4
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997
ACCESSIONS S41346
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
#accession S41346
##molecule_type genomic RNA
##residues_ 1-118 #label VAN
##cross-references EMBL:229449
##experimental_source genotype 1, N4
CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS capsid protein; core protein; polyprotein
FEATURE 1-118
#product core protein #status predicted #label MAT
SUMMARY #length 118 #checksum 81

Query Match 99.2%; Score 775; DB 9; Length 118;
Best Local Similarity 89.9%; Pred. No. 2.23e-106;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnpkqrktrkntnrpqdvkfgggqivgvyllprgprlgvratrktsersqprg 60
|||||
Qy 1 MSTNPKQRKTRKNTNRPRXXXXPGGGIVGGVLLPRGPRXGVRATRKTSERSQPRG 60

Db 61 rrcqpkarrpgrtwagpywplyngcgwagwllsprgrpswqp 109
|||||
Qy 61 RRQPIKAXRKGSRWAQPGYPMPLYNGEGCGWALLSPRGRPNWGP 109

RESULT 7
ENTRY #type complete
TITLE genome polyprotein - hepatitis C virus (strain H)
CONTAINS capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS1; nonstructural protein NS2; nonstructural protein NS3; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
ORGANISM #formal_name hepatitis C virus
#note host Homo sapiens (man)
DATE 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Feb-1997
ACCESSIONS A36814; A41546
REFERENCE A36814
#authors Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
#submission submitted to GenBank, July 1992
#description Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates.

#accession A36814
##molecule_type genomic RNA
##residues_ 1-3011 #label INC
##cross-references GB:M67463
REFERENCE A41546
#authors Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:10292-10296
#title Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates.

#cross-references MUID:92052256
#contents annotation
#note neither amino acid nor nucleotide sequence is given
CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS capsid protein; envelope protein; glycoprotein; nonstructural protein; polyprotein; transmembrane protein

FEATURE 1-115
#product capsid protein C #status predicted #label CPC\
116-191 #product envelope protein M #status predicted #label EPV\
192-389 #product major envelope protein E #status predicted #label MEE\
390-729 #product nonstructural protein NS1 #status predicted #label NS1\
730-1006 #product nonstructural protein NS2 #status predicted #label NS2\
1007-1615 #product nonstructural protein NS3 #status predicted #label NS3\
1616-1862 #product nonstructural protein NS4a #status predicted #label N4A\
1863-2013 #product nonstructural protein NS4b #status predicted #label N4B\
2014-3011 #product nonstructural protein NS5 #status predicted #label NS5\
#product nonstructural protein NS5 #status predicted #label NS5\
#product nonstructural protein NS5 #status predicted #label NS5

```

196, 209, 234, 305,
325, 417, 423, 430,
448, 476, 532, 540,
556, 576, 623, 645,
1213, 1255, 2041,
2240, 2364, 2789
#binding site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY      #length 3011 #molecular-weight 327143 #checksum 438
Query Match   99.2%; Score 775; DB 4; Length 3011;
Best Local Similarity 89.9%; Pred. No. 2.23e-106;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db    1 metnpharktkntnrpqdvkfpqggqlvgvylpprqprlgavratrkterssqrg 60
      |||||
Qy    1 MSTNPKPQRKTKRNTNRPPXXXXXPGGGQIVGCGVYLLPRRGPRGXGVRAKTRKTSERSQPRG 60
      |||||

Db    61 rrrpiakrrpggrtwagpygypwplyngccgwagwllsprgerpswp 109
      |||||
Qy    61 RRRIPIKAEXGRSQAOCYPWPPLYNGECGWAXWLLSPRGSRNNWGP 109

```

RESULT	8
ENTRY	S41365
TITLE	genome polyprotein - hepatitis C virus (genotype 4, N4) (fragment)
CONTAINS	core protein
ORGANISM	#formal_name hepatitis C virus
#variety	genotype 4, N4
DATE	19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997
ACCESSIONS	S41365
REFERENCE	S41361
#authors	van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission	submitted to the EMBL Data Library, January 1994
#description	Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
#accession	S41365

```

#molecule_type genomic RNA
#residues 1-114 #label VAN
#cross-references EMBL:Z29468
#experimental_source genotype 4, N4
CLASSIFICATION superfamily hepatitis C virus genome polyprotein
KEYWORDS capsid protein; core protein; polyprotein
FEATURE
1-114 #product core protein #status predicted #label
SUMMARY
#length 114 #checksum 8888

```

Query Match 98.5%; Score 769; DB 9; Length 114;
Best Local Similarity 89.9%; Pred. No. 2.22e-105;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```

Db      1  mstnpkprkktkntnrmpmdvfkpgggqivggvyliprrgprlgvratrktsersqprg  60
      |||||
Ov      1  MSTNPKPQRTKTKNTNRMPXXXXXPGGGQIVGGVYLIAPRRGPRXGVVTRKTSERSOPRG  60

```

Db

61 rrrqipkarqrsegrswagpgypwplygncgcwgawlllsgsrpsrpswp 109
||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Ov

61 RRRIIPKARXREGRSWAOPGYDPWPLYGNCGCWAXWLISPRSRPNWP 109

RESULT	ENTRY	TITLE
9	S41362	#type fragment genome polypeptide - hepatitis C virus (genotype 4, N1) (fragment)

```

CONTAINS      core protein
ORGANISM     #formal_name hepatitis C virus
              #variety      genotype 4, N1
DATE         19-May-1994  #sequence_revision 26-Jul-1996  #text_change
              16-Feb-1997
ACCESSIONS   S41362
REFERENCE     S41341
              van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
              submitted to the EMBL Data Library, January 1994
              Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
              #accession    S41362
              #molecule_type genomic RNA
              #residues     1-114  ##label VAN
              #cross-references EMBL:229465
              #experimental_source genotype 4, N1
CLASSIFICATION  #superfamily hepatitis C virus genome polypeptide
KEYWORDS        capsid protein; core protein; polypeptide
FEATURE         #product core protein #status predicted #label MAT
                1-114
SUMMARY         #length 114 #checksum 8846

```

Query Match 98.5%; Score 769; DB 9; Length 114;
Best Local Similarity 89.9%; Pred. No. 2.22e-105;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db	1	mstnkpqr	tkrntnrr	pmdivk	fpagq	ivggv	llpr	gprlg	vrat	rkts	ers	sqr	60
Qv	1	MSTNKPQR <td>TKRNTNRR <td>PMDIVK <td>FPAGQ <td>IVGGV <td>LLPR <td>GPRLG <td>VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td></td></td></td></td></td></td></td>	TKRNTNRR <td>PMDIVK <td>FPAGQ <td>IVGGV <td>LLPR <td>GPRLG <td>VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td></td></td></td></td></td></td>	PMDIVK <td>FPAGQ <td>IVGGV <td>LLPR <td>GPRLG <td>VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td></td></td></td></td></td>	FPAGQ <td>IVGGV <td>LLPR <td>GPRLG <td>VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td></td></td></td></td>	IVGGV <td>LLPR <td>GPRLG <td>VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td></td></td></td>	LLPR <td>GPRLG <td>VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td></td></td>	GPRLG <td>VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td></td>	VRAT <td>RKTS <td>ERS <td>SQR <td>60</td> </td></td></td>	RKTS <td>ERS <td>SQR <td>60</td> </td></td>	ERS <td>SQR <td>60</td> </td>	SQR <td>60</td>	60

```

Db      61  rrqpikarqpegrswagpgypwplynecgawgllsprgsrswgp 109
        ||||| : ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oa      61  PROPtKAYRxEGRSwaOPGYDpWPLYNecGCWAXWLLSPGSRRWGP 109

```

RESULT	10
ENTRY	JQ1584
TITLE	#type fragment genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

CONTAINS	core protein C; envelope protein EI; envelope protein E2; nonstructural protein NSI
ORGANISM	#normal name hepatitis C virus
DATE	17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 03-May-1994

ACCESSIONS	REFERENCE
JQ1584	Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J. #authors
JQ1584	J. Gen. Virol. (1992) 73:1521-1525 #journal
	Cloning and sequencing of the structural region and #title
	expression of putative core gene of hepatitis C virus from a British case of chronic sporadic hepatitis.

#accession Q01584
##molecule_type genomic RNA
##residues 1-640 **##label** KUM
CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS core protein; envelope protein; glycoprotein; nonstructural protein; polyprotein

```

FEATURE
1-191      #product core protein C #status predicted #label CPC\
192-389    #product envelope protein E1 #status predicted #label
           EE\
390-640    #product envelope protein E2 and nonstructural protein
           NS1 #status predicted #label EE2\

```

196, 209, 234, 305,
417, 430, 448, 476,


```

SUMMARY          #length 115   #checksum 9764

Query Match      97.6%; Score 762; DB 9; Length 115;
Best Local Similarity 89.0%; Pred. No. 3.23e-104;
Matches 97; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Db 1 mstnpgqrktkntnrpqdvpfgggvgvylwprgprlqvratrktsersqrg 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1 MSTNPKPQRKTKNTNRPRXXXXXPGGGVYLVLRGPRGVTRTKTTSRSQPRG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 rrcpiakrrpcrwtacpywpvllyngcgcwgawllspgrtspwqp 109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1 RRPIPKAAXRKGSRWAQGPYPWPLYNGEGCGMAXLLSPGRSPPNWGP 109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
ENTRY genome polyprotein - hepatitis C virus (genotype 1, N11)
TITLE (fragment)
CONTAINS core protein
ORGANISM #formal name hepatitis C virus
#variety genotype 1, N11
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change
16-Feb-1997

ACCESSIONS S41343
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
#accession S41343
#molecule_type genomic RNA
#residues 1-115 #label VAN
#cross-references EMBL:Z29446
#experimental_source genotype 1, N11
CLASSIFICATION #superfamily hepatitis C virus genome polypeptide
KEYWORDS capsid protein; core protein; polypeptide
FEATURE
1-115 #product core protein #status predicted #label MAT
SUMMARY          #length 115   #checksum 9697

```

```
Search completed: Mon Feb 23 11:02:03 1998
Job time : 56 secs.
```

Feb 23 20:47

/home/fuller/feb1098/US-08-836-075A-1.rge

1

Release 2.1D John F. Collins, BioComputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MParch_nm n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Mon Feb 23 20:43:03 1998; MasPar time 435.69 Seconds
Tabular output not generated.
Title: >US-08-836-075A-1
Description: (1-327) from US08836075A.seq
N.A. Sequence: 317
Comp: 1 ATGAGCAGCAATCTTAACCC.....CTGGCCCAATTTGGGCCCC 327
TACTCTGCTAGGATTTGG.....GAGCGGCTTACCCCGGG

Scoring table:
TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 430261 seqs, 710217276 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:
emb1-new7
1: BCT 2: FUN 3: GEN 4: HTG1 5: HTG2 6: HTG3 7: HTG4 8: HUM1
9: HUM2 10: HUM3 11: INV1 12: INV2 13: ORG 14: NAM 15: VPT
16: PLN 17: PRO1 18: PRO2 19: ROD 20: SYN 21: UNC 22: VIR
genbank101
23: BCT1 24: BCT2 25: BCT3 26: BCT4 27: BCT5 28: BCT6 29: BCT7
30: BCT8 31: BCT9 32: BCT10 33: BCT11 34: BCT12 35: BCT13
36: GEN1 37: GEN2 38: GEN3 39: GEN4 40: GEN5 41: GEN6 42: HTG1
43: HTG2 44: HTG3 45: HTG4 46: HTG5 47: INV1 48: INV2 49: INV3
50: INV4 51: INV5 52: INV6 53: INV7 54: INV8 55: INV9 56: INV10
57: INV11 58: INV12 59: MAM1 60: MAM2 61: MAM3 62: VRT1
63: VRT2 64: VRT3 65: VRT4 66: PAT1 67: PAT2 68: PAT3 69: PAT4
70: PAT5 71: PAT6 72: PAT7 73: PHG 74: PLN1 75: PLN2 76: PLN3
77: PLN4 78: PLN5 79: PLN6 80: PLN7 81: PLN8 82: PLN9 83: PLN10
84: PLN11 85: PLN12 86: PRI1 87: PRI2 88: PRI3 89: PRI4
90: PRI5 91: PRI6 92: PRI7 93: PRI8 94: PRI9 95: PRI10
96: PRI11 97: PRI12 98: PRI13 99: PRI14 100: PRI15 101: PRI16
102: PRI17 103: ROD1 104: ROD2 105: ROD3 106: ROD4 107: ROD5
108: ROD6 109: ROD7 110: ROD8 111: ROD9 112: STR 113: SYN
114: UNA
genbank101
115: VRL1 116: VRL2 117: VRL3 118: VRL4 119: VRL5 120: VRL6
121: VRL7 122: VRL8 123: VRL9 124: VRL10 125: VRL11
genbank-new7
126: BCT 127: GEN 128: HTG1 129: HTG2 130: INV 131: MAM

(TM)

Feb 23 20:47

/home/fuller/feb1098/US-08-836-075A-1.rge

2

Database: 132:VRT 133:PHG 134:PLN1 135:PLN2 136:PRI1 137:PRI2
138:ROD 139:SYN 140:UNA 141:VRL
u-emb151_101
142:part1 143:part2

Statistics: Mean 9.880; Variance 4.750; scale 2.080

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
1	317	100.0	327	143	A50346			Sequence 1 from Paten	7.68e-244
2	291	91.8	506	117	HCVHN3			Hepatitis C virus cor	4.45e-221
3	285	89.9	573	116	HCU10204			Hepatitis C virus iso	7.81e-216
4	285	89.9	573	116	HCU10193			Hepatitis C virus iso	7.81e-216
5	285	89.9	577	69	E05083			DNA encoding a part o	7.81e-216
6	285	89.9	5400	122	HPCCGENOM			Hepatitis C China vir	7.81e-216
7	285	89.9	9410	122	HPCK1S1			Hepatitis C virus (st	7.81e-216
8	285	89.9	9410	122	HPCK1R1			Hepatitis C virus (st	7.81e-216
9	284	89.6	1613	122	HPCKNS1SPC			Hepatitis C virus cor	5.84e-215
10	283	89.3	327	143	A50350			Sequence 5 from Paten	4.36e-214
11	283	89.3	346	117	HCVGRIN5			Hepatitis C virus (ge	4.36e-214
12	283	89.3	573	116	HCU10209			Hepatitis C virus iso	4.36e-214
13	283	89.3	1863	122	HPCHCJ2			Hepatitis C virus (HC	4.36e-214
14	283	89.3	9410	122	HPCK1R2			Hepatitis C virus (st	4.36e-214
15	283	89.3	9410	122	HPCK1S2			Hepatitis C virus (st	4.36e-214
16	283	89.3	9427	122	HPCKRNA			Hepatitis C virus str	4.36e-214
17	282	89.0	326	122	HPCHCV006			Hepatitis C virus gen	3.26e-213
18	282	89.0	326	122	HPCHCV007			Hepatitis C virus gen	3.26e-213
19	282	89.0	9535	115	D85516			Hepatitis C virus gen	3.26e-213
20	281	88.6	506	117	HCVHB1			Hepatitis C virus cor	2.43e-212
21	281	88.6	506	117	HCVHN2			Hepatitis C virus cor	2.43e-212
22	281	88.6	573	116	HCU10225			Hepatitis C virus iso	2.43e-212
23	281	88.6	573	116	HCU10203			Hepatitis C virus iso	2.43e-212
24	281	88.6	573	116	HCU10189			Hepatitis C virus iso	2.43e-212
25	281	88.6	573	116	HCU10202			Hepatitis C virus iso	2.43e-212
26	281	88.6	848	117	HCU63376			Hepatitis C virus cor	2.43e-212
27	281	88.6	906	68	E03572			cDNA sequence encodin	2.43e-212
28	281	88.6	1551	122	HPCTSTRUCTB			Hepatitis C virus cor	2.43e-212
29	281	88.6	1620	122	HPCE2COR			Hepatitis C virus cor	2.43e-212
30	281	88.6	1880	122	HPCTSTRJ4			Hepatitis C virus RNA	2.43e-212
31	281	88.6	2540	68	E04260			cDNA encoding a part	2.43e-212
32	281	88.6	2540	69	E04805			cDNA to 5'-terminal r	2.43e-212
33	281	88.6	2540	69	E07391			cDNA encoding chimpan	2.43e-212
34	281	88.6	4129	122	HPCTOLYPRO			Hepatitis C virus (HC	2.43e-212
35	281	88.6	9391	68	E04421			DNA encoding peptide	2.43e-212
36	281	88.6	9391	69	E07579			Gene coding Hepatitis	2.43e-212
37	281	88.6	9414	122	HPCKNKCD5			Hepatitis C virus mRN	2.43e-212
38	281	88.6	9415	116	HCU16362			Hepatitis C virus pol	2.43e-212
39	281	88.6	9440	124	S62220			polyprotein [Hepati	2.43e-212
40	281	88.6	9448	122	HPCK483			Hepatitis C virus gen	2.43e-212
41	281	88.6	9456	122	HPCKRNA			Hepatitis C virus RNA	2.43e-212
42	280	88.3	326	122	HPCHCV010			Hepatitis C virus gen	1.82e-211
43	280	88.3	326	122	HPCHCV2F			Hepatitis C virus (is	1.82e-211
44	280	88.3	326	122	HPCHCV077			Hepatitis C virus gen	1.82e-211
45	280	88.3	2433	143	A48711			Sequence 49 from Paten	1.82e-211

ALIGNMENTS

RESULT 1
ID A50346 standard; DNA; UNC; 327 BP.
AC A50346;
NI e1039185
DT 07-MAR-1997 (Rel. 51, Created)
DT 07-MAR-1997 (Rel. 51, Last updated, Version 1)
DE Sequence 1 from Patent W09613590.
KW
OS unidentified
OC unclassified.
RN [1]
RP 1-327
RA Maertens G., Stuyver L.;
RT "NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE
AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS";
RL Patent number W09613590-A/1, 09-MAY-1996.
RL INNOGENETICS NV (BE).
CC Other publication AU 3844095 960523
FH Key Location/Qualifiers
FH source 1..327
FH /organism="unidentified"
SQ Sequence 327 BP; 56 A; 102 C; 104 G; 51 T; 14 other;
Query Match 100.0%; Score 317; DB 143; Length 327;
Best Local Similarity 100.0%; Pred. No. 7.68e-244;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 atgagcagcaatcctaaacctcaaaagaaacaaacgtaacacacacacccgcgcctcaak 60
Qy 1 ATGAGCAGCAATCCTTAACCTCAAGAAACCAACAGGTAAACACCAACCGCGCCCTCAK 60
Db 61 gsgtgnnnnnnccgggtggcggtcagatcggtgtgagttacacctgttgcgcgcagg 120
Qy 61 GSGGTNNNNNNCCGGGTGGCGGTCAAGTCTGTGGTGGAGCTTTACCTGTGGCGCGCAGG 120
Db 121 ggcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 180
Qy 121 GGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
Db 181 agcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
Qy 181 AGCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240
Db 241 tatecltggccctctatggcaatgaggggtgcgggtggcggtggcggtggcggtggcggtggcggt 300
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGT 300
Db 301 cgcggtctcggcgaattgggcccc 327
Qy 301 CGCGGTCTCGGCGCAATTGGGCCCC 327

RESULT 2
LOCUS HCV/HN3 506-bp DNA VRL 29-APR-1994
DEFINITION Hepatitis C virus core gene, HN3 isolate.
ACCESSION X76410
NID 6434947
KEYWORDS core; core protein.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.

REFERENCE 1 (bases 1 to 506)
AUTHORS Qu, D.
TITLE Direct Submission
JOURNAL SUBMITTED (22-NOV-1993) to the EMBL/GenBank/DBJ databases. D. Qu,
INSERM Unite 271, U. de Recherche sur les Hepatites, le Sida et
les Retrovirus Humains, 151 cours Albert Thomas, F 69424 Lyon Cedex
03, FRANCE
REFERENCE 2 (bases 1 to 506)
AUTHORS Qu, D., Hantz, O., Gouy, M., Vitvitski, L., Li, J.S., Berby, F.,
Tong, S.P. and Trepo, C.
TITLE Heterogeneity of hepatitis C virus genotypes in France
J. Gen. Virol. 75 (Pt 5), 1063-1070 (1994)
MEDLINE 94231157
FEATURES
source Location/Qualifiers
1..506
/organism="Hepatitis C virus"
/isolate="HN3"
misc_feature 29..506
/notes="region corresponding to the core protein of the
polyprotein of HCV"
BASE COUNT 86 a 166 c 166 g 88 t
ORIGIN
Query Match 91.8%; Score 291; DB 117; Length 506;
Best Local Similarity 91.7%; Pred. No. 4.45e-221;
Matches 300; Conservative 4; Mismatches 23; Indels 0; Gaps 0;
Db 29 atgagcagcaatcctaaacctcaaaagaaacaaacgtaacacacacccgcgcgcagcag 88
Qy 1 ATGAGCAGCAATCCTTAACCTCAAGAAACCAACAGGTAAACACCAACCGCGCCCTCAK 60
Db 89 gacgtcgaagtcctccggcggtggtcagatcggtgtgagttacacctgttgcgcgcagcag 148
Qy 61 GSGGTNNNNNNCCGGGTGGCGGTCAAGTCTGTGGTGGAGCTTTACCTGTGGCGCGCAGG 120
Db 149 ggcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 208
Qy 121 GGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
Db 209 agcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 268
Qy 181 AGCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240
Db 269 tacccttggccctctatggcaatgaggggtgcgggtggcggtggcggtggcggtggcggtggcggt 328
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGT 300
Db 329 cgtggtctcggcctagtgtgggcccc 355
Qy 301 CGCGGTCTCGGCGCAATTGGGCCCC 327
RESULT 3
LOCUS HCV10204 573 bp RNA VRL 26-AUG-1994
DEFINITION Hepatitis C virus isolate P10 core protein gene.
ACCESSION U10204
NID g532392
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
REFERENCE 1 (bases)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.

23
-14
9
318
327 - 977b

FNWATKIKLTP IPAASQLDLSNFWVAGYSGDVYHLSRPRMFMLCILLISVGV

```
GYLLPNR"
330..902
/product="core protein"
903..1478
/product="E1"
1479..2756
/product="E2"
1479..1559
/notes="hypervariable region (HVR)"
2757..3407
/product="NS2"
3408..5300
/product="NS3"
5301..5462
/product="NS4A"
5463..6245
/product="NS4B"
6246..7586
/product="NS5A"
6954..7073
/notes="interferon sensitivity determining region (ISDR)"
7587..9359
/product="NS5B"
1904 a 2846 c 2691 g 1969 t
```

BASE COUNT
ORIGIN

```
Query Match 89.9%; Score 285; DB 122; Length 9410;
Best Local Similarity 90.8%; Pred. No. 7.81e-216;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 330 atgagcagcaatcctaaccctcaaaagaaacaaacgttaacacacacccgcgcacag 389
|||||
Qy 1 ATGACACCAATCTCTAACTCAAGAAACCAACACGCTAACACCAACCGCCCTCAK 60
|||||

Db 390 gacgtcaagtcccgccggtggtcagatcgttgtagagttacctgtgcccgcagg 449
|||||
Qy 61 GCGCTNNNNNNCCGGCTGCGCGTCAGATCGTTGTCGAGTTTACTCTGTCGCGCAGG 120
|||||

Db 450 gccccaggttggtgtgcgcgcactaggaaactccgagcgtgcgaacctcgtgga 509
|||||
Qy 121 GCGCCACAGNNGGCTGCGCGCGACTAGGAGACTTCGACGCGGTCAACACTCGTGC 180
|||||

Db 510 aggcgacaacctatccccagagctcgccgcgcaggcggcggcctggcctcagccggg 569
|||||
Qy 181 AGCGGACACCTATCCCCACGCTCGCGGCGGAGCGGCGGCGGCTCGTGGCGTCAGCGCGG 240
|||||

Db 570 taccctggccctctatgcgaatgagccttaggtggcagcagtgctcctgcaacc 629
|||||
Qy 241 TATCCTTGCGCCCTCTATGCGCAATGAGGCTCGCGGTCGCGGCGGCGGTCGCTGTCGCC 300
|||||

Db 630 cgcgcgtcccgccctagttggggcccc 656
|||||
Qy 301 CGCGGCTCGCGCCCAATTGGGGGCCCC 327
|||||
```

```
RESULT 8
LOCUS HPCK1R1 9410 bp as-RNA VRL 02-FEB-1997
DEFINITION Hepatitis C virus (strain HCV-1b, clone HCV-K1-R1), complete genome
sequence.
ACCESSION D50480
NID g1030706
KEYWORDS polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
NS5b; envelope protein; non-structural protein;
```

```
interferon-sensitive; interferon-resistant; IFN-sensitive;
IFN-resistant; ISDR; interferon sensitivity determining region;
HVR; hypervariable region.
SOURCE Hepatitis C virus (strain:HCV-1b) cDNA to genomic RNA,
clone:HCV-K1-R1.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
REFERENCE 1 (sites)
AUTHORS Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
TITLE Comparison of full-length sequences of interferon-sensitive and
resistant hepatitis C virus 1b. Sensitivity to interferon is
conferred by amino acid substitutions in the NS5A region
J. Clin. Invest. 96 (1), 224-230 (1995)
JOURNAL 95340824
MEDLINE 2 (bases 1 to 9410)
REFERENCE Enomoto,N.
AUTHORS Unpublished (1995)
JOURNAL 3 (bases 1 to 9410)
REFERENCE Enomoto,N.
AUTHORS Direct Submission
TITLE Submitted (08-MAY-1995) to the DDBJ/EMBL/GenBank databases.
JOURNAL Nobuyuki Enomoto, Tokyo Medical and Dental University, Second
Department of Internal Medicine; 1-5-45 Yushima, Bunkyo-ku, Tokyo
113, Japan (E-mail:PXN04522@niftyserve.or.jp,
Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)
```

```
FEATURES
source
1..9410
/organism="Hepatitis C virus"
/notes="IFN-resistant HCV from patient 1"
/strain="HCV-1b"
/clones="HCV-K1-R1"
330..9362
/codon_start=1
/products="polyprotein"
/db_xref="PID:d1009706"
/db_xref="PID:g1814085"
/translation="MSTNPQRKTKNTNRPDQVKPGGGQIVGGYVLLPRGPRL
GVRATKTSRQRRRQIPKARRPEGMAAQGTPWFLYNGELGAGHLLSPRG
SRPQGPDPRRSRNLGKVIDITCGFADLMGYPIVCGAPLGGARALAHGVILED
GVNATGILPGCSFSLFLLALLSLITPASAVEVCNAGLYHVNDSCNSSIVIEAQD
MIMTPGCVPCVRENNSSRCWALPTLAARNASIPPTTIRHVDLIVGAAFCSAMV
VGDLCGSVFVLSQFLTFSPPRYETVDQNCSTYPGHVSGHRMAMMMNSPTTALV
SOLLIPQAVDMVAGAHMGVLGAYYSWGNWAKVLIIVMLLFAGVDGNTTGAQ
GRTVSRLLTSITPGASQNTQLINTNGSHINRTALNCNDSLQATGFIALFYARRENS
GCPQRASCRS IDKEAQQMGPIYAEHSDQRPYQWRYAPRCPTGPIPAQVCGPVIC
FTPSVVGTTDRFCAPTYNMGATDVLANTTPYQGMFCCTMNGTGFCTKCGG
PPCNIIGCAGNNTLCTDTCFKHPEATYAKCGSGPMLTPRCIVDYPVRLMHPCTVNF
SIFKRVMYGGVGHRLTAACNWRGRCDEDRSELSPLLSTTEMQILPCSTFTLL
PALSTGLIHLQNTVDVQVLYGVGSVAVSFAMKMEYVLLFLLLADARVACALWMLL
VAQAAALENIIVLNNAALAGVHGLSFLVEFCAMV IKGLVPGAAYLVGWPLLL
LLIALPPAYAMDRMAASCGGAVF IGIALLTSLPYKVF LAKLILWLIQYLI TRAEAQ
LQWNPINVRGRDAIILLTCAPHEIFDITKIVLAILGCLP LMYLQAGTAKYPTDFVR
AQQVIRECTLVBKVGGGVQVQMAFMKLAELKGT SVYDHLTPLRDWAHTGLRDLAVAVE
PVFSDMEIKIITWGGDTACGDII MGILPVSARRGRELILGPADSLEGQGRLLAPIT
AYSQQTREGLFGCIITSLTGRDKNGVEGEVQVVSSTATQSF LQTCVNGVCWTVYHGASK
TLAGKPTITQMYNTVDQLVGMPPAPCARSLTCTCGGSDLYLVTRHADVIPVRRRG
DSRGSLLSRPISYIKGSCSGCLPCPSCHVVGIFRAAVCTRGVAKAVDFVPSMETT
MRSVPVFDNSPPAPQPTQVAHLHAPTGSKSTKVPAAQAQGYKVLVLPVSVTATL
GFAYMSKAGHVDNIRITGVTITTTGAPIT YTYSTYKGLADGCGSGGAYDIIMCDECHS
IDSTSILGCTVLDQAEACARLVLATATPPTCVSTVPHNIEETALSNTGELIPFYCK
AIP IETIKGRHLIFCHSKKKCKDELAKLSGLGLNNAVYYRGLDVSVIPASGVVVVA
TDALMTGTFGDFSDVDCNTCVTQTVDFSLDPTFTTETTTVPQDAVSRSQRRTGRG
```


NID	g532402	
KEYWORDS		
SOURCE	Hepatitis C virus.	
ORGANISM	Hepatitis C virus	
	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis C-like viruses.	
REFERENCE	1 (sites)	
AUTHORS	Bukh,J., Purcell,R.H. and Miller,R.H.	
TITLE	Sequence analysis of the core gene of 14 hepatitis C virus genotypes	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 91, 8239-8243 (1994)	
MEDLINE	94336721	
REFERENCE	2 (bases 1 to 573)	
AUTHORS	Bukh,J., Purcell,R.H. and Miller,R.H.	
TITLE	Sequence analysis of the 5' noncoding region of hepatitis C virus	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)	
MEDLINE	92279243	
REFERENCE	3 (bases 1 to 573)	
AUTHORS	Bukh,J., Purcell,R.H. and Miller,R.H.	
TITLE	At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (17), 8234-8238 (1993)	
MEDLINE	93376778	
REFERENCE	4 (bases 1 to 573)	
AUTHORS	Bukh,J.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-MAY-1994) Bukh J., National Institutes of Health, Hepatitis Viruses Section, 9000 Rockville Pike, Bethesda, MD 20892, USA	
FEATURES		
source	Location/Qualifiers	
	1..573	
	/organism="Hepatitis C virus"	
	/isolate="S45"	
	/specific_host="Homo sapiens"	
	1..573	
	/gene="C"	
	/standard_name="core"	
	/function="nucleocapsid protein"	
	/evidence=experimental	
	/product="core protein"	
	1..573	
gene	/gene="C"	
CDS	1..573	
	/gene="C"	
	/codon_start=1	
	/product="polyprotein"	
	/db_xref="PID:g532403"	
	/translation="MSTNPKPQRTKNTNRPDVKFPGGQIVGGVILPFRPRRL	
	GYRATKTSERSQPRGRQIPKARPEGRAWQPGHPPLYGNEGIGWAGWLLSPRG	
	SRP SWGFTPPRRSRNLGVKVIDTTCGFADLMGYIPLVGAP LGGAARALAHGVRIED	
	GVNATGNLPGCSFIFLLALLSGITIPASA"	
BASE COUNT	95 a 189 c 177 g 112 t	
ORIGIN		
Query Match	89.3%; Score 283; DB 116; Length 573;	
Best Local Similarity	90.5%; Pred. No. 4,366-214;	
Matches	296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;	
Ddb	1 atgagcagcaatcctaaactcaagagcaaacacaaactgaacacaaacccgccccacag 60	
Qy	1 ATGACGACGAATCCTAAACCTCAAAAGAAAAACCAACGTAACACCAACGCCGCCCTCAK 60	
Ddb	61 gacgtcaagttccggggtggcggtcagatcgtttggtaggttacctgtgcgcgcagg 120	

Qy	61	GGSGTNUNNNCCGGTGGCGGTGACATCGTTGGTGGAGCTTACCTGTTGGCGCGGAGG	120
Db	121	ggccccagtttaggtgtgqgcgcgaactagaagaacttcqagcaggtcacaaactcgtgga	180
Qy	121	GGCCCAAGGNNGGGTGTGGCGCGCACTAGGAAGACTTCGACGGGTACAACTCGTGGC	180
Db	181	cggcgacaacatccccaaagctgcgcggccagggcaggcaggcctggccagcccggg	240
Qy	181	AGGCGACAGCCTATCCCAAGGCTGTCGGYCCGAGGCGAGGTCTGTGGGCTCAGCCGGG	240
Db	241	catccttggccctctatggcaatgagggcttggggtgggcaggatggctcctgtcacc	300
Qy	241	TATCTGTGTCGGCTCTATGCAATGAGGCTGCGGTGCGGGTGGCTGTCTGTGCCCC	300
Db	301	cgtggctccggcctagttggggcccc	327
Qy	301	CGGGCTCTGGGCCCAATGGGGCCCC	327
RESULT	13		
LOCUS	HPCHCJ2	1863 bp ss-RNA	VRL 05-AUG-1994
DEFINITION	Hepatitis C virus (HCV) genomic RNA for structural protein, partial sequence.		
ACCESSION	D10074		
NCBI	g221588		
KEYWORDS	structural protein.		
SOURCE	Hepatitis C virus (isolate:HC-J2) cDNA to genomic RNA.		
ORGANISM	Hepatitis C virus		
REFERENCE	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis C-like viruses.		
AUTHORS	Okamoto,H., Kuri,K., Okada,S., Yamamoto,K., Lizuka,H., Tanaka,T., Fukuda,S., Teuda,F. and Mishiro,S.		
TITLE	Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes		
JOURNAL	Virology 188 (1), 331-341 (1992)		
MEDLINE	92230232		
REFERENCE	2 (bases 1 to 1863)		
AUTHORS	Okamoto,H.		
JOURNAL	Unpublished (1991)		
COMMENT	Data kindly submitted in computer readable form by: Hiroaki Okamoto		
FEATURES	Immunology Division Jichi Medical School Kawachi-gun Tochigi 329-04 Japan Phone: 0285-44-2111 x3334 Fax: 0285-44-1557.		
source	Location/Qualifiers 1..1863 /organism="Hepatitis C virus" /isolate="HC-J2" 325..>1863 /codon_start=1 /product="structural protein" /db_xref="PID:d1001434" /db_xref="PID:g221589"		
CDS	translation="MSTNPKQKTRNTRRRPDVKFPGGQIVGCVLIPRGRPL GVRAIRKTSRQSRGRRQIPKARPEGRMAQPGTPPLYNEGELGWAQGLLSPRG SRPSWGPTDPRRRSRLGKVIDTUTCGFADLMGYIPLVGAPLGGNARALAHGVRLD SYNATVGNLPCCSFIFLALLSLCITIPASAYEVRNVSQVHTVNDCSNSSIIVYEAD LIMEATGCPVCRENNSSRCWVALTPTLAARTS IPTTTIRRHVDLLVGAAATFCSAAY		

VGDLCGSFVLVSQFTSPRRHETLQDNCSTYPGHLSGHRMAWDMWMMNSPTTALVW
SOLLRIPTOTVDMVTGAHCVLAGLAYYSVMGNWAKVLIYVLLFAGVDGTHVTGAT
GHFTSGIASLPLPGASQKIQIINTNGSHHNRFLANCNDSLNTGFLAALFYTHKFNAS
GCPERLASCRSDGFDGDMGPTTYTEPQDSQDKPYCHMYAPQRCSWSAADVCGPVYC
FTRPSP*

BASE COUNT 344 a 574 c 539 g 406 t

ORIGIN

Query Match 89.3%; Score 283; DB 122; Length 1863;
Best Local Similarity 90.5%; Pred. No. 4.36e-214;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

Db 325 atgagcagcaatcctaaacctcaagaagaaacaaacgtaacacacacacgctgcgcacag 384
|||||
|||||

Qy 1 ATGAGCAGCAATCTTAACCTCAAAAGAAAACCAACAGCTAACACACCGCGCCCTCAK 60
|||||

Db 385 gacgtcaagttcccgggcgcggtcagatcggttggtgsgagttacctgttgcgcgcag 444
|:|||
|||||

Qy 61 GCGCTNNNNNNCCGGTGGCGGTGACATCGTGTGAGATTACCTGTTGCGCGCAGG 120
|||||

Db 445 gccccaggttggtggtgcgcgcgactaggaagacttcgagcgttcgcaacctcgtgga 504
|||||
|||||

Qy 121 GGGCCCAAGNNNGGTGTGGCGCGACTAGGAAGACTTCCGAGCGGTCAACACTCTGTGGC 180
|||||

Db 505 agggcgaacactatccccaggtcgcgcgcgactccgagcgcagcgttcgagcgcgcg 564
|||||
|||||

Qy 181 AGCGCAGACGCTATCCCAAGCGCTGCGGCGCGAGCGGCTCTGCGGCTCAGCCCGGG 240
|||||

Db 565 taacctggccctctatgcaatgaggtcaggtgaggtgaggtgaggtcaggtcaggtcagcc 624
||
|||||

Qy 241 TATCTTGGCCCTCTATGGAATGAGGCTCGCGGCTGGCGCGGCTGCTCTGCTGCC 300
|||||

Db 625 cgtggctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 651
||
|||||

Qy 301 CGCGGCTCTGCGCCCAATGGGGGCC 327

RESULT 14
LOCUS HPCK1R2 9410 bp ss-RNA VRL 02-FEB-1997
DEFINITION Hepatitis C virus (strain HCV-1b, clone HCV-K1-R2), complete genome
sequence.
D50481

ACCESSION g1030705
NID polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
KEYWORDS NS5b; envelope protein; non-structural protein;
interferon-sensitive; interferon-resistant; IFN-sensitive;
IFN-resistant; ISDR; interferon sensitivity determining region;
HVR; hypervariable region.

SOURCE Hepatitis C virus (strain:HCV-1b) cDNA to genomic RNA,
clone:HCV-K1-R2.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.

REFERENCE 1 (bases)
AUTHORS Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
TITLE Comparison of full-length sequences of interferon-sensitive and
resistant hepatitis C virus 1b. Sensitivity to interferon is
conferred by amino acid substitutions in the NS5A region
J. Clin. Invest. 96 (1), 224-230 (1995)

JOURNAL 95340824

MEDLINE 2 (bases 1 to 9410)

AUTHORS Enomoto,N.

JOURNAL Unpublished (1995)

REFERENCE 3 (bases 1 to 9410)
AUTHORS Enomoto,N.
TITLE Direct Submission
JOURNAL Nobuyuki Enomoto, Tokyo Medical and Dental University, Second
Department of Internal Medicine; 1-5-45 Yushima, Bunkyo-ku, Tokyo
113, Japan (E-mail:PXN04522@niftyserve.or.jp,
Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)

FEATURES Location/Qualifiers
source l..9410
/organism="Hepatitis C virus"
/note="IFN-resistant HCV from patient 2"
/strain="HCV-1b"
/clones="HCV-K1-R2"
330..9362
/codon_start=1
/product="polyprotein"
/db_xref="PID:d1009707"
/db_xref="PID:g1814086"

CDs
/translation="MSINPQRKTKNTRRPQDVKEPGGGQIVGGYLLPRGRRL
GVRAFRKTSRSPGRQRQIPKARQPEGRMAQPCYPWPLYNGECMGAGWLLSPRG
SRPSKGPDPRRSRNKLVIDTLTCGFADLMCYIPIVGLAPLGASRALAHGVRVLED
GVNATGNLPGCSFIFLALLSCLITPASAEVVRNASGVYHVTNDCNSASIVYEAD
MIHTPGCPVCREANSRCWALPTTLAARNSSVPTTIRRHVDLIVGAAATCSAMY
VGDLCGSFVLVSQFTSPRRHETVQDNCSTYPGHLSGHRMAWDMWMMNSPTTALVW
SOLLRIPTQAVMDMVAGAHGCVLAGLAYYSVMGNWAKVLIYVLLFAGVDGTHVTGAT
SHITGGLSIFAPGAQKIQIINTNGSHHNRFLANCNDSLNTGFLAALFYTHKFNAS
GCPERIASCRSDAFAQMGPIITYAEPGSSDQRPYCHMYAPRPGIIPASEVCGPVYC
FTRPSPVVTGTRDSCAPTYSWGENETDVLILNANTRAPQGNWECCTMNGTGTCTKCGG
PPCNIIGVGNNTLTCTDCKRKHPEATYTKGSGPWLTPRCWVDYPIYRLMHPCTVNF
TIFKVRMYGGVHRLNAQNRTRCEDRDELSELSPLLSLTFEQLIPLGSEFTL
PVSLGLIHLHONIVDQVLYGVGSVSVFKWEYIILLLFLLADARVACALWMLL
IAQAEALENULVNAASVAGAGHGLFIVFFCAAYIKGKLVPGCAAYAFYGVWPLLL
LILALPPRAYAMDREMAASCGGAEVGLALLTLSPHYKVIARLIWLIQYFIYRAEC
LQWVIPPVNRGRDAIILITCALHPELIFSITKILVALIPLVLAOGITKMPYFVR
AQGLIRACHLVKRVAGHYVQMAFKLAALITGTYVDHLTPLDQWHTGLRLDANVE
PVVSDMETKIIITWGAADTACGDIILGLAVSARRKEIFLGPADSLGSGWRLLAPIT
AYSQGTGLIGCIIITSLTRGDKNOVEGQVOWSTATOSFIATCTGTCVMTVHGASK
TLAGKPCPITQMYTNVDQDLVQKQAPSGASLTPCTCGSSDLYLVTRHADYIPVRERG
DGRGSLSPRVSYLKGSGGLPCPGHAGVGFRAAVCTRGVAKAVDFIPVESMET
MESPVFTDMSPPAPQTFQVAHLHAPTGSGSKNKPVEYAAQYKVLVLPVSAATL
GFGAYMSKAGVDNIRITGRTITTGAPITTYTGKFLADGGCGGAYDIIICDECHS
TDSTSLIGITGVLDQAGARLVVLATATPPGSVTPHPNIEVALNTEGTEIPFYK
AIPETIKGRHLIFCHSKKCDLAKSLGLGINAVAYRGIDVSVIPASGDVVVA
TDALMTGTFDSDVDICTVTQVDFSLDPTETIETTVODAVSRORRGTGRG
RAGYRVTPGERSGMDSVLCEDACGAWELPFAETSVRLRAYLNTPLGVPWQ
DLHEFWESVFTGLTHDAHGLSQTQKQDNFPIYVAQVTCARQAQAPPSWDMWKC
LIRLPLTHGPTLLYHAGVQNEVILTHPIKYIMACMSADLEVVTSTWLVGGVLA
ALAAVCLTGSVIVGRIILSKPAILPDREVLVYEFDEMEECASHLPYIEQCMQIAE
QFKQALIGLIQTATQAEAAVWVESKQDALETFWAKHMNFISGTYLAGLSLTPGN
PAIASLMAFTASITSLTQHTLFTNLLGQWAAQLAPPRAASAFVAGIAGAAGS1
GLGKVLVDILAGYAGVAGALVAFKWSGDMPTEDLVNLLPAILSPCALVVGWCAA
ILRRHVCGEGAVQMMRLIFASGHNVSPHYVVPESDAARVQTQLSSLTITQLK
PLIKHINEDCSTPCSLRDVDMIC*TVLSDFKTMQLSRVPLRPVPLFSCQRYK
GYWPGDGIQGTTCGGAQITGHVKNQSMRIVGPKTCNTWHTGTFINAYTTGCTPSP
APNYSRALWRVAEEYVEVTRVQDGHVYVGTMTDNIAKQCPQAPAEFFKELDQVRLR
YAPASRLLRDEVTFQVGLNQVDFGSLPCEPEPDVALVLSMTDPSHITAEATKRLR
ARGSPSSASSQSLSAPSIKATCTTHDSDADLLEANLLWROEMGNIITVESEN
KVVILDSFEPLEADEDESVSLPRLIRKSKFPAMPVWARPDPNPPLIESHKDPDY
VPPVHCGPLPTKAPPIPPRKRTVYTESVSSALAEATKTFGSSSSAAQST
ATAPPDQSSGDGDSVESYSNPELEGEPDPLSDGSMSTVSEEGADWCCSWS
YTWGALITPCAESKLPINALSNLLRHNNVYATTSRSASQRQKVTDFDLVD
DHYRDLVLEKAKASTYKARLLSVEEAKLTPPHSARSKFGYGAADVRNLSRAINHI
RSWKMLQEDTETPTDITIMAKNEVFCVQPEKGGKAPRLIVFPDILGVRVCEKALYD


```
VVSTLPQAVMGASGYFQYSPQCRVEE LVNAWKSCKPMGFAYDTRCFDSTVTENDIRV
EES1YQCCDLAPARQALRSILTERLYIGPLTNSKQNGCYRRCRASGLVTTSCGNLT
TCYLKATACRAKQLQDCTHLVCGDDLVVIGESAGTQEDAAASLVNFTETMRISAPPG
DPQPEYDLELITSCSSNVSVANDASGRVYVYLTTRDPTTP LARAAMETARHTPNSWL
GNLWYAPT LWABMLMTHFFSILLAQOLEKALDCQIYGACYSIEPLDIPQIQRHL
GLSASLHUSYSPCEINRVASCLBKLVPPLEVRHRRARSVRAKLLSQGGRAALCGKYL
FNWVRTKMLTTPIPAASQDLJSGMFVAGYSGGBIYHVSVRNRPWFNMCILLLSVGV
GYLLPNR"
330..902      /product="core protein"
903..1478     /product="E1"
1479..2756   /product="E2"
1479..1559   /note="hypervariable region (HVR)"
2757..3407   /product="NS2"
3408..5300   /product="NS3"
5301..5462   /product="NS4A"
5463..6245   /product="NS4B"
6246..7586   /product="NSA"
6954..7073   /note="interferon sensitivity determining region (ISDR)"
7587..9359   /product="NS5B"
BASE COUNT   1889 a 2853 c 2700 g 1968 t
ORIGIN
```

```
Query Match      89.3%; Score 283; DB 122; Length 9410;
Best Local Similarity 90.5%; Pred. No. 4,36e-214;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

Db 330 atgacacgaatcctaaacctcaagaagaaacacaaacotaaacacacccgccgcacag 389
| : ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Qy 1 ATGAGCAGCAATCTCTAAACCTCAAGAAACCAACCAACGTAACCAACCCGCCCTCAK 60

Db 390 gacgtcaagttcccgggcggtggtcagatcgttgtagagtttacctgttgcgcgcagg 449
| : ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GGGTNNNNNNCCGGGTGGCGGTGACATCGTTGCTGAGTTTACCTGTTCGCCGCCGAGG 120

Db 450 ggcgccaggttggtgtgctgcgcgaacttaggaagactccgagcggtgcgaacctcgtgga 509
| : ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 GGCCCCAGGNGGCTGTGCGCGACTAGGAAGACTTCCGAGCGGTCACACCTCGTGGC 180

Db 510 aggcacaaacctatcccaaggttcgcagccgagcgagcgagcggtcggctcagcccgagg 569
| : ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 AGCGGACAGCCTATCCCAAGGCTCGYCGGCGCCAGCGGACAGCTCCTGGGCTCAGCCGGG 240

Db 570 taccttgccctctatgccaatgagggcatgggggtgggcaggaatgctcgtgcaacc 629
| : ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 TATCTTGGCCCTCTATGGCAATAGGGCTGCGGGTGGCGGGTGGCTGCTGTCCTCC 300

Db 630 cgcgcgtctcgcctcagttggggccccc 656
| : ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 CCGGGCTCTCGGCCCAATTGGGGCCCC 327
```

```
RESULT 15
LOCUS HPC152 9410 bp ss-RNA VRL 02-FEB-1997
```

```
DEFINITION Hepatitis C virus (strain HCV-1b, clone HCV-K1-S2), complete genome
sequence.
ACCESSION D50485
NID g1030704
KEYWORDS polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
NS5B; envelope protein; non-structural protein;
interferon-sensitive; interferon-resistant; IFN-sensitive;
IFN-resistant; ISDR; interferon sensitivity determining region;
HVR; hypervariable region.
SOURCE Hepatitis C virus (strain:HCV-1b) cDNA to genomic RNA,
clone:JCVC-K1-S2.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
REFERENCE 1 (sites)
AUTHORS Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
TITLE Comparison of full-length sequences of interferon-sensitive and
resistant hepatitis C virus 1b. Sensitivity to interferon is
conferred by amino acid substitutions in the NS5A region
J. Clin. Invest. 96 (1), 224-230 (1995)
JOURNAL 95340824
MEDLINE 2 (bases 1 to 9410)
REFERENCE Enomoto,N.
AUTHORS Unpublished (1995)
JOURNAL 3 (bases 1 to 9410)
REFERENCE Enomoto,N.
AUTHORS Direct Submission
TITLE Submitted (08-MAY-1995) to the DDBJ/EMBL/GenBank databases.
JOURNAL Nobuyuki Enomoto, Tokyo Medical and Dental University, Second
Department of Internal Medicine; 1-5-45 Yushima, Bunkyo-ku, Tokyo
113, Japan (E-mail:PXN04522@niftyserve.or.jp,
Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)
```

FEATURES

```
source
1..9410
/organism="Hepatitis C virus"
/note="IFN-sensitive HCV from patient 2"
/strain="HCV-1b"
/clones="HCV-K1-S2"
330..9362
/codon_start=1
/product="polyprotein"
/db_xref="PID:g1814089"
/translation="MSTNPKPQKTKRNNRPPQDKFPGGQIVGGVYLLPRGRPL
GVATKTSERSQPRGRQPIPKARQPEGRWAQPGYWPVLYGNECMGAGNLLSPRG
SRPSWGTDPRRRSNLGKVIDTLTCGFDLMGYIPVVGAP LGGASRALAHGVRVLED
GVNATGNLFGGSFISFLALLSLCTIPASAEVRNASGVYHTNDCSNASIVYEAD
MTHMTPCVPVCREANSSRCWVALLTPLAARNSSVPTTIRRHVDLVYGAATCSAMY
VGBLCGSVFLVSQLETFSPRRHETVQDNCISYIPGSHVSGHRMAMMMNWSPTTALVQ
SOLLIRIPQAVMDMVAAGHWGLAGLAYSVMVGNWAKVLIWLLFLAGVDGRTVTGGAG
GHITQRLASLFTGCAQRIQLVNTNGSMHINRTALNCNDSINTGFLAALFYTHRENS
GPTERASCRSIDAFAQWGPITYAEPSSDQRPYCHHYAPRPGGIVPASEVCGPYVC
FTPSVWVGTDRS GAPTSGWENETDVLNNTTRAPQGNMFCCTMMNGTFTKTCGG
PPCNI GGNGNLTLCPTDCFRKPEATYTKCGSGPWLTPRCWVDYPIYRLMHPCTWNF
TIFKRYMVGGEHRLNAACNWRGECGDLDRSLSPLLSLTETQIILPCSFYTL
PVLSTGLIHQNIYDVQYLYGVGSVAEVIKWEYILILLFLLIADARVACLWMLL
IAOEAEMENLVINAASVAGAGFLSFLVFFCAAWYIKGLVPGAAVAFYGVWPLL
LLALPPRANVMDREMAASCGGAVFGVGLALLTLSPHYKVLARLIMWLQYETIRAEAC
LQWIPPIPNVRRGRDAIILLTCAHPILFISITKILNAILGRLMWLQAGITKMPYFVR
AOGILRACMLRVKAGGHYQMAFMKLAALTGTYYVDHLTPLDQWAHAGLRDLAWE
PVVFSDMETKILTWGADTACGDIITGLIYASABRGKEIFLCPADSLEGSGWRLAPIT
AYSQDTRGLLGCIITSLTGDRKNQVEGEVQVNSTATQSFATCTNGVCWTVYHGASK
TLASGKPIQWMTYTNVDQDLVGNQAPSGARSLTPCTCGSSDLYLVTRHADYIPVRRG
```


Result No.	Score	Query		DB ID	Description	Type	Pred. No.
		Match	Length				
1	317	100.0	327	25	T27937	Hepatitis C virus	5,348-189
2	285	89.9	573	21	T16620	Hepatitis C virus	1,166-167
3	285	89.9	573	21	T16621	Hepatitis C virus	1,166-167
4	283	89.3	327	25	T27939	Hepatitis C virus	2,496-166
5	283	89.3	570	3	Q21747	Hepatitis C virus	2,496-166
6	283	89.3	573	21	T16617	Hepatitis C virus	2,496-166

CC The sequences T27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
 CC This sequence represents nucleotides 1-310 from the HCV type 1d isolate
 CC BNL1.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides R96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 CC Sequence 327 BP; 56 A; 102 C; 104 G; 51 T;

Query Match 100.0%; Score 317; DB 25; Length 327;
 Best Local Similarity 100.0%; Pred. No. 5,34e-189;
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 atgagcacgaatcctaaacctcaaaagaaacacaaacgtacacacacgcgcctctcak 60
 |||||
 Qy 1 ATGAGCAGCAATCCTTAACCTCAAGAAACCAACACGTAACACACCGCGCCTCAK 60
 |||||
 Db 61 gqgtgtnnnnnnccgggtggcggtcgagatcggtggagtttaacctgttgcgcgcagg 120
 |||||
 Qy 61 GSGTNNNNNNCCGGGTGGCGGTCAAGATCGTGGTGGAGTTTACCTGTTGCCGCGAGG 120
 |||||
 Db 121 ggcacacagmggtgtgcgcgcactaggaagacttcagagcggtcacaacctcgtggc 180
 |||||
 Qy 121 GCGCCACAGNNNGGTGTCGCGCGACTAGGAAGACTTCGAGCGGTCAACAACCTCGTGGC 180
 |||||
 Db 181 aggcagacagctatccccaaagctcgcgcgcagggcgcgcaggtcggctcagcccgagg 240
 |||||
 Qy 181 AGGCAGACAGCTATCCCCAAGGCTCGCGYCCGCGGCGAGGCTCCTGGGCTCAGCCCGGG 240
 |||||
 Db 241 tatcttgcccctctatggcaatgagggctggcggtggcgggntggctcctgtccccc 300
 |||||
 Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGGTGGCGGGNTGGCTCCTGTCCCCC 300
 |||||
 Db 301 cgcgctctcgcgcgcattggggccccc 327
 |||||
 Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 2
 ID T16620 standard; cDNA; 573 BP.
 AC T16620;
 DT 01-OCT-1996 (first entry)
 DE Hepatitis C virus isolate P10 core protein gene.
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT CDS 1..573
 FT /*tag= a
 FT /product= core_protein
 FT /note= "does not contain stop codon"
 PN W09605315-A2.
 PD 22-FEB-1996.
 PF 15-AUG-1995; U10398.

PR 15-AUG-1994; US-290665.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI; 96-139709/14.
 DR P-PSDB; R92946.
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 PS Claim 3; Page 156; 340pp; English.
 CC T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV
 CC isolates. The isolated sequences are useful for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
 CC used in the prevention of HCV infection.
 CC Sequence 573 BP; 96 A; 185 C; 179 G; 113 T;

Query Match 89.9%; Score 285; DB 21; Length 573;
 Best Local Similarity 90.8%; Pred. No. 1.16e-167;
 Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 1 atgagcacgaatcctaaacctcaaaagaaacacaaacgtacacacacgcgcgcacag 60
 |||||
 Qy 1 ATGAGCAGCAATCCTTAACCTCAAGAAACCAACACGTAACACACCGCGCCTCAK 60
 |||||
 Db 61 gacgtcaagttcccgcggtggtcgagatcggtggagtttaacctgttgcgcgcagg 120
 |||||
 Qy 61 GSGTNNNNNNCCGGGTGGCGGTCAAGATCGTGGTGGAGTTTACCTGTTGCCGCGAGG 120
 |||||
 Db 121 ggcacacagmggtgtgcgcgcactaggaagacttcagagcggtcacaacctcgtggc 180
 |||||
 Qy 121 GCGCCACAGNNNGGTGTCGCGCGACTAGGAAGACTTCGAGCGGTCAACAACCTCGTGGC 180
 |||||
 Db 181 aggcagacacatccccaaagctcgcgcgcagggcgcgcaggtcggctcagcccgagg 240
 |||||
 Qy 181 AGGCAGACAGCTATCCCCAAGGCTCGCGYCCGCGGCGAGGCTCCTGGGCTCAGCCCGGG 240
 |||||
 Db 241 tatcttgcccctctatggcaatgagggctggcggtggcgggntggctcctgtccccc 300
 |||||
 Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGGTGGCGGGNTGGCTCCTGTCCCCC 300
 |||||
 Db 301 cgcgctctcgcgcgcattggggccccc 327
 |||||
 Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 3
 ID T16621 standard; cDNA; 573 BP.
 AC T16621;
 DT 01-OCT-1996 (first entry)
 DE Hepatitis C virus isolate DK1 core protein gene.
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT CDS 1..573
 FT /*tag= a
 FT /product= core_protein
 FT /note= "does not contain stop codon"
 PN W09605315-A2.
 PD 22-FEB-1996.
 PF 15-AUG-1995; U10398.

PR	15-AUG-1994; US-290665.	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PA	(USSH) US SEC DEPT HEALTH.	
PI	Bukh J, Miller RH, Purcell RH;	
DR	WPI; 96-139709/14.	
DR	P-PSDB; R92947.	
PT	DNA and amino acid sequence of HCV envelope 1 and core proteins -	
PT	used to determine HCV genotype and as vaccines against HCV infection	
PS	Claim 3; Page 156-157; 340pp; English.	
CC	T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV	
CC	isolates. The isolated sequences are useful for the prodn. of primers	
CC	useful for detecting the presence of HCV in a sample, the primers	
CC	are also useful for HCV genotyping. Proteins encoded by the cDNAs	
CC	can be used in vaccines for immunising against HCV infection. The	
CC	proteins may also be used to detect antibodies against HCV in serum,	
CC	saliva, lymphocytes or other mononuclear cells. The antibodies may be	
CC	used in the prevention of HCV infection.	
SQ	Sequence 573 BP; 98 A; 185 C; 177 G; 113 T;	
	Query Match 89.9%; Score 285; DB 21; Length 573;	
	Best Local Similarity 90.8%; Pred. No. 1.16e-167;	
	Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;	
Db	1 atgagcagcaatcctaaccctcaagaagaaacccaaacgtaacaccacgcgcgcacag 60	
Qy	1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACGTAACACCAACGCGCGCCTCAK 60	
Db	61 gacgtcaagtcccgcggtggttcagatcgttgggtgagtttacctgttcgcgcgagg 120	
Qy	61 GCGGTNNNNNNCCGGGTGGCGGTCAAGTCTGTGGTGAAGTTTACCTGTTCGCGCGAGG 120	
Db	121 ggcgccaggttggtgtgcgcgcgaactaggaagactccgcagcggtcgcaacctcgtgga 180	
Qy	121 GGCCCCAGAGNNGGTGTGGCGCGACTAGGAAGACTTCCGAGGGGTCAACCTCGTGGC 180	
Db	181 aggcgaacaacctatcccgaagctgcggccgagggcagggcctgggtcgaaccggg 240	
Qy	181 AGGCGACAGCCTATATCCCAAGGCTCGYCGYCCGAGGGCAGGCTCTGGGCTCAGCCGGG 240	
Db	241 tacccttgccctctatggcaatgaggcgatgggttggcagaggtgacctcgtcaccc 300	
Qy	241 TATCCTTGGCCCCCTCTATGGCAATGAGGGCTCGGGGTGGGGGGTGGCTCTCTCTCCCC 300	
Db	301 cgcgcctctcgccctagttggggcccc 327	
Qy	301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327	
RESULT	4	
ID	T27939 standard; DNA; 327 BP.	
AC	T27939;	
DT	10-MAR-1997 (first entry)	
DE	Hepatitis C virus type 1d isolate BNL2 bases 1-310.	
KW	Hepatitis C virus; subtype; polymerase chain reaction; amplification;	
KW	PCR; primer; probe; antibody; infection; ss.	
OS	Hepatitis C virus.	
PN	W09613590-A2.	
PD	09-MAY-1996.	
PF	23-OCT-1995; E04155.	
PR	21-OCT-1994; EP-870166.	
PR	28-JUN-1995; EP-870076.	
PA	(INNO-) INNOGENETICS NV.	
PI	Maertens G, Stuyver L;	
DR	WPI; 96-251460/25.	

DR P-PSDB; R96528.

PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type

PT - used to develop probes and primers for new sub:types and vaccines

PT to prevent and treat infection

PS Claim 6; Fig 3; 150pp; English.

CC The sequences T7937-T72989 represent novel sequences isolated from

CC hepatitis C virus subtypes different from subtypes la-c, 2a-d, 3a-f,

CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,

CC 3q, 4k-m, 7a-z or types 9, 10 or 11. The sequences corresp. to the 5'

CC untranslated region (UR), the Core/EI, NS4 or NS5B regions of the genome.

CC This sequence represents nucleotides 1-310 from the HCV type 1d isolate

CC BNL12.

CC The new HCV types were isolated from patients with chronic HCV from the

CC Benelux countries, France, Cameroon and Vietnam, because of their

CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR

CC amplified, cloned and genotyped. The 5'UR, Core/EI and NS5B regions were

CC sequenced either directly or partially and used to classify the new

CC viruses into (sub)types based on comparison with known sequences.

CC The sequences were used to generate the peptides R96424-R96524. The

CC sequences can also be used to synthesise probes and primers for the

CC detection of HCV in a sample. The polypeptides can be used to detect

CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.

SQ Sequence 327 BP; 62 A; 105 C; 103 G; 54 T;

Query Match 89.3%; Score 283; DB 25; Length 327;

Best Local Similarity 90.5%; Pred. No. 2.49e-166;

Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

Db 1 atgagcagcaatcctaataaccataaagaaacaaacgaacacacccgcgcacag 60

Qy 1 ATGAGCAGCAATCTTAACACTCAAA GAAAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db 61 gagctcaagttccgggtgtggtcagatcgttggtggagttacctgttgcgcgaggg 120

Qy 61 GGSGTNNNNNNCCGGGTGGCGGTCAGATCGCTTGTGTGAGTTTACCTGTTCGCGCGAGS 120

Db 121 ggcgccagttgggtgtgagcgaccaggaagacttcgagcggtcgagcctcgtgcac 180

Qy 121 GGC GCCCAGNNNGGTGTGCGCGGCACTAGGAAGACTTCCGAGCGGTCAACACTCTGGC 180

Db 181 aggcagacgcctatctcaagctcgccagtcogagtggcagmccctgggtcagccaggg 240

Qy 181 AGGCGACACCTATCCCAAGCTGCGGTCGCGGTCGAGGGGAGGCTCTTGGGCTACGCCGGG 240

Db 241 catccttgcccctctatggcaatgagggctgcgatggcgggatggctcctgtcccc 300

Qy 241 TATCCTTTGGCCCTCTATTGGCAATGAGGGCTGCGGGTGGGCGGNTGGCTCTGTCCCC 300

Db 301 cgcggtctcgcccgagttgggccc 327

Qy 301 CGCGGCTCTCGGCCAATTGGGCCCC 327

RESULT 5

ID Q21747 standard; DNA; 570 BP.

AC Q21747;

DE 22-JUN-1992 (first entry)

DT HCV clone Th cDNA.

KW HCV1; ss.

OS Hepatitis C virus.

PN W09202642-A.

PD 20-FEB-1992.

PF 12-AUG-1991; U05728.

PR 10-AUG-1990; IJS-566209.

PA (CHIR-) CHIRON CORP.
 PI Houghton M, Choo QL, Kuo G, Weiner AJ, Urdea MS, Irvine BD;
 PI Kolberg JA;
 DR WPI; 92-080094/10.
 PT Reagents for isolating, amplifying and detecting HCV
 PT polynucleotide(s) - used to monitor spread of blood-borne non-a,
 PT non-b hepatitis virus infection and screen blood samples for
 PT virus
 PS Disclosure; Fig 2; 67pp; English.
 CC The sequences represented in Q21746-50 are DNA consensus
 CC sequences for five different HCV isolates from different
 CC geographic locations (Japan and U.S.).
 SQ Sequence 570 BP; 94 A; 183 C; 177 G; 116 T;
 Query Match 89.3%; Score 283; DB 3; Length 570;
 Best Local Similarity 90.5%; Pred. No. 2.49e-166;
 Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
 Db 1 atgagcagcaatcctaaacctcaaaagacaaacaaacgtaacacacacgctcgccacag 60
 Qy 1 ATGAGCAGCAATCCTAAACCTCAAGAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
 Db 61 gacgtcaagttcccggtgagcgctcagatcggtgaggttacttctgtgcccgcagg 120
 Qy 61 GCGGTNNNNNNCGGGTGGCGGTGAGATCGTTGCTGAGTTTACTCTTGGCGCGCAGG 120
 Db 121 ggcctcaggttgggtgtgctgcgcagaggaagacttccgagcggtcgcaacctcgaggt 180
 Qy 121 GGCCCCAGGNNGGGTGTGCGCGGCACTAGGAAGACTTCGAGCGGTCAACAACCTCGTGGC 180
 Db 181 agactcagcctatccccagcagctcgcccgagggcaggagactgggtcagccggg 240
 Qy 181 AGGGGACAGCCTATCCCGAGGCTGCGGCGGCGAGGCGAGGCTCTGGGCTCAGCCGGG 240
 Db 241 tatcttgcccctctatggcaatgagggctgaggtggcgaggtggtctctctctccc 300
 Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGGTGGCGGGGNTGGCTCTGTCCCCC 300
 Db 301 cgtggctcgcgcctagctggggcccc 327
 Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 6
 ID T16617 standard; cDNA; 573 BP.
 AC T16617;
 DT 01-OCT-1996 (first entry)
 DE Hepatitis C virus isolate S45 core protein gene.
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT CDS 1..573
 FT /*tag= a
 FT /product= core_protein
 FT /note= "does not contain stop codon"
 PN W09605315-A2.
 PD 22-FEB-1996.
 PF 15-AUG-1995; U10398.
 PR 15-AUG-1994; US-290665.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI; 96-139709/14.

DR P-PSDB; R92943.
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 PS Claim 3; Page 154; 340pp; English.
 CC T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV
 CC isolates. The isolated sequences are useful for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
 CC used in the prevention of HCV infection.
 SQ Sequence 573 BP; 95 A; 189 C; 177 G; 112 T;
 Query Match 89.3%; Score 283; DB 21; Length 573;
 Best Local Similarity 90.5%; Pred. No. 2.49e-166;
 Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
 Db 1 atgagcagcaatcctaaacctcaaaagacaaacaaacgtaacacacacgctcgccacag 60
 Qy 1 ATGAGCAGCAATCCTAAACCTCAAGAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
 Db 61 gacgtcaagttcccggtgagcgctcagatcggtgaggttacttctgtgcccgcagg 120
 Qy 61 GCGGTNNNNNNCGGGTGGCGGTGAGATCGTTGCTGAGTTTACTCTTGGCGCGCAGG 120
 Db 121 ggcctcaggttgggtgtgctgcgcagaggaagacttccgagcggtcgcaacctcgaggt 180
 Qy 121 GGCCCCAGGNNGGGTGTGCGCGGCACTAGGAAGACTTCGAGCGGTCAACAACCTCGTGGC 180
 Db 181 agactcagcctatccccagcagctcgcccgagggcaggagactgggtcagccggg 240
 Qy 181 AGGGGACAGCCTATCCCGAGGCTGCGGCGGCGAGGCGAGGCTCTGGGCTCAGCCGGG 240
 Db 241 tatcttgcccctctatggcaatgagggctgaggtggcgaggtggtctctctctccc 300
 Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGGTGGCGGGGNTGGCTCTGTCCCCC 300
 Db 301 cgtggctcgcgcctagctggggcccc 327
 Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 7
 ID Q38959 standard; cDNA; 9391 BP.
 AC Q38959;
 DT 29-JUL-1993 (first entry)
 DE Hepatitis C virus gene.
 KW HCV; detection; antigen; vaccine; recombinant; ss.
 OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT CDS 305..9349
 FT /*tag= a
 FT /product= core_protein
 FT /note= "does not contain stop codon"
 PN J05068563-A.
 PD 23-MAR-1993.
 PF 17-JUL-1991; 203884.
 PR 17-JUL-1991; JP-203884.
 PA (KAGA) KAGAKU OYOBI KESSEI RYOHO.
 DR WPI; 93-130639/16.
 DR P-PSDB; R35207.
 PT Nucleotide sequence encoding hepatitis C virus polypeptide - is
 PT useful for detecting HCV infection, esp. in Japan, and as vaccine
 PT against HCV
 PS Claim 1; Page 6-17; 17pp; Japanese.

CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
 CC used in the prevention of HCV infection.

SQ Sequence 573 BP; 96 A; 184 C; 177 G; 116 T;

Query Match 88.6%; Score 281; DB 21; Length 573;
Best Local Similarity 90.2%; Pred. No. 5,34e-165;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 atgagcagcaatcctaaacctaagaataaaccaaacgtaaacacaacccgcgcacag 60
|||||
Qv 1 A'GAGCGCAATCTCTAAACCTCAAGAGAAAACCAACGTAACACCAACCGCCGCCCTCAK 60

Db 61 gacgtcaagttccccggcggtcgcagatcgttggtggagttaacctgtagccgcgcagg 120
| : | ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Ov 61 GSGGTNNNRNCCGGTGGCGGTCAGATCGTTGGTGAGTTACTGTTGCCGC GCAGG 120

Db 121 gaccccgagttgggtatgcgcgcgactagaagacttcaggcggctcgcaacctcgtgga 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Gv 121 GACCCGAGCGNNGCTGGCGCGCACTAGGAAGACTTCGAGCGGTCAACCTCTGCGC 180

Db 181 aggcgaacactatccccaggctgccgcggccgagggtagggtcctgggtcagccggg 240

[illegible]

DB 301 cgcggctcccggcctagtggggccc 327
||||||| ||||| | ||||| |||||
327

RESULT 11

AC T16622;
DT 01-OCT-1996 (first entry)
DE Hepatitis C virus isolate T10 core protein gene.

KW	hepatitis; ss.
OS	Hepatitis C virus.
FH	Key
	Location/Qualifiers

```
FT /*tag= a
FT /product= core_protein
FT /note= "does not contain stop codon"
```

PD 22-FEB-1996.
PF 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.

PA (USSH) US SEC DEPT HEALTH.
PI Bukh J, Miller RH, Purcell RH;
DR WPI: 96-139709/14.

PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
ps Claim 3. Page 157. 340nm. English
p-PSDBJ: K92346.

CC isolates. The isolated sequences are useful for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV

Query Match 88.6%; Score 281; DB 21; Length 573;
Best Local Similarity 90.2%; Pred. No. 5,34e-165;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 atgagcagcaatcctaaacctaagaataaaccaaacgtaaacacaacccgcgcacag 60
|||||
Qv 1 A'GAGCGCAATCTCTAAACCTCAAGAGAAAACCAACGTAACACCAACCGCCGCCCTCAK 60

Db 61 gacgtcaagttccccggcggtcgcagatcgttggtggagttaacctgtagccgcgcagg 120
| : | ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Ov 61 GSGGTNNNRNNCCGGTGGCGGTCAGATCGTTGGTGAGTTACTGTTGCCCGCGCAGG 120

Db 121 gaccccgagtgtgggtatgcgcgcgactagaagacttcgcagcgtcgcaacctcgtgga 180

Ov 121 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 180

121 gaccccgacgcnngcctgcgcgcgcgactaggagacattccgacgcgtcacacctctctggc 180

Db 181 aggcgaacactatccccaggctgccgcggccgagggtagggtcctgggtcagccggg 240

Db

241 tacccttgccocttatggcaacgagggttgggtggcaggatggctctgtcccc 300
|||||
|||||

G

241 ttatgcctggcgccgcttttcctcccattgaactgtagggctgacgtgcccc 300
|||||
|||||

DB 301 cgcggctccccggcctagtggggcccc 327
||||||| ||||| | ||||| |||||
327

RESULT 11

AC T16622;
DT 01-OCT-1996 (first entry)
DE Hepatitis C virus isolate T10 core protein gene.

KW	hepatitis; ss.
OS	Hepatitis C virus.
FH	Key
	Location/Qualifiers

```
FT /*tag= a
FT /product= core_protein
FT /note= "does not contain stop codon"
```

PD 22-FEB-1996.
PF 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.

PA (USSH) US SEC DEPT HEALTH.
PI Bukh J, Miller RH, Purcell RH;
DR WPI: 96-139709/14.

PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
ps Claim 3. Page 157. 340nm. English
p-PSDBJ: K92346.

CC isolates. The isolated sequences are useful for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV

CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
 CC used in the prevention of HCV infection.

SQ Sequence 573 BP; 96 A; 184 G; 177 G; 116 T;

Query Match 88.6%; Score 281; DB 21; Length 573;
 Best Local Similarity 90.2%; Pred. No. 5.34e-165;
 Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 atgagcagaatcctaaacctcaagaagaaacacaaacgtaacacacacgcccagcag 60
 |||||
 Qy 1 ATGAGCAGCATCTTAACCTCAAGAAACCAACGCTAACACACACCGCGCCTCAK 60

Db 61 gacgtcaagtcccgggcggtggtcgatcggtggagttacctgttgcgcgcagg 120
 |||||
 Qy 61 GCGCTNNNNNNCCGGGTGGCGGTGCATCGTGTGGTGAGTTTACCTGTTCGCGCGCAGG 120

Db 121 ggcgccaggttgggtatgcgcgcactagaagacttcgacggtcgcaacctcgtgga 180
 |||||
 Qy 121 GCGCCGCGGNNNGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCAACACCTCGTGGC 180

Db 181 aggcgaacctatcccaagctcgcgcgcagcgcgcaggttagggcttgcgcgcagg 240
 |||||
 Qy 181 AGCGCAGACGCTATCCCAAGGCTCGYCGYCGGACGCGGAGGCTCTGGCTCAGCGCGG 240

Db 241 tacccttgccctctatggcaacgaggttgggtgggcaggatggctctgtcaccc 300
 |||||
 Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTCGCGGTGGGCGGNTGGCTCTGTGCCCC 300

Db 301 cgcggtcccggtctagtggggcccc 327
 |||||
 Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 14
 ID T16625 standard; cDNA; 573 BP.
 AC T16625;
 DT 01-OCT-1996 (first entry)
 DE Hepatitis C virus isolate IND8 core protein gene.
 KW HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis; ss.
 OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT CDS 1..573
 FT /*tag= a
 FT /product= core_protein
 FT /note= *does not contain stop codon*
 PN W09605315-A2.
 PD 22-FEB-1996.
 PF 15-AUG-1995; U10398.
 PR 15-AUG-1994; US-290665.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI; 96-139709/14.
 DR P-FSD8; R92951.
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 PS Claim 3; Page 139; 340pp; English.
 CC T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV
 CC isolates. The isolated sequences are useful for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
 CC used in the prevention of HCV infection.

SQ Sequence 573 BP; 97 A; 185 G; 177 G; 114 T;

Query Match 88.6%; Score 281; DB 21; Length 573;
 Best Local Similarity 90.2%; Pred. No. 5.34e-165;
 Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 atgagcagaatcctaaacctcaagaagaaacacaaacgtaacacacacgcccagcag 60
 |||||
 Qy 1 ATGAGCAGCATCTTAACCTCAAGAAACCAACGCTAACACCAACCGCGCCTCAK 60

Db 61 gacgtcaagtcccgggcggtggtccagatcggtggagttacctgttgcgcgcagg 120
 |||||
 Qy 61 GCGCTNNNNNNCCGGGTGGCGGTGCAGATCGTGTGGTGAGTTTACCTGTTCGCGCGCAGG 120

Db 121 ggcgccaggttgggtatgcgcgcactagaagacttcgacggtcgcaacctcgtgga 180
 |||||
 Qy 121 GCGCCGCGGNNNGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCAACACCTCGTGGC 180

Db 181 aggcgaacctatcccaagctcgcgcgcagcgcgcaggttagggcttgcgcgcagg 240
 |||||
 Qy 181 AGCGCAGACGCTATCCCAAGGCTCGYCGYCGGACGCGGAGGCTCTGGCTCAGCGCGG 240

Db 241 cacccttgccctctatggcaatgaggttgggtgggcaggatggctctgtcaccc 300
 |||||
 Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTCGCGGTGGGCGGNTGGCTCTGTCCCC 300

Db 301 cgcggtctcgcgcctagtggggcccc 327
 |||||
 Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 15
 ID Q44921 standard; DNA; 686 BP.
 AC Q44921;
 DT 04-OCT-1994 (first entry)
 DE Hepatitis C virus 5'-untranslated region and part of core region.
 KW Hepatitis C virus; HCV; non-A, non-B hepatitis virus; NANBHV;
 KW antisense oligonucleotide; translation inhibition; therapy;
 KW 5'-untranslated region; 5'-UTR; hybridisation target; ss.
 OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT 5'UTR 1..341
 FT /*tag= a 342..686
 FT CDS
 FT /*tag= b
 FT /product= Core_protein
 FT /note= *partial coding region*
 PN W09405813-A.
 PD 17-MAR-1994.
 PF 10-SEP-1993; J01293.
 PR 10-SEP-1992; US-945289.
 PR 14-APR-1993; JP-087195.
 PA (KAGA) CHEMO SERO THERAPEUTIC RES INST.
 PA (ISIS-) ISIS PHARM INC.
 PA (MOCH) MOCHIDA PHARM CO LTD.
 PI Anderson KP, Eto I, Furukawa S, Hamada F, Hanecak RC;
 PI Hoshiko K, Nakatake H, Nishihara T, Nozaki C;
 DR WPI; 94-101217/12.
 PT Anti-sense oligonucleotide(s) complementary to hepatitis C viral
 PT genome - useful for inhibiting HCV replication, to treat related

Job time : 64 secs.

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 114:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: DK1
SQ Sequence 573 BP; 98 A; 185 C; 177 G; 113 T; 0 other;

Query Match 89.9%; Score 285; DB 13; Length 573;
Best Local Similarity 90.8%; Pred. No. 2.94e-183;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGCTACACACCGCCGCCACAG 60
Qy 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGCTACACACCGCCGCCCTCAK 60

Db 61 GACGTCAAGTTCGCGCGGTGTCAGATCGTTGCGAGTTTACCTGTGCGCGCAGG 120
Qy 61 GCGTNNNNNNCCGGTGGCGGTGATCGTTGTCGCTTACCTGTGCGCGCAGG 120

Db 121 GCGCCGAGTTGGTGTGCGCGGCTAGGAGACTTCCGAGCGGTGCGCAACCTCGTGA 180
Qy 121 GCGCCGAGNNGGTGTGCGCGGCTAGGAGACTTCCGAGCGGTGCGCAACCTCGTGGC 180

Db 181 AGCGACACCTATCCCAAGGCTGCGCGGCGCGGCGGCGGCTGCGCTCAGCCCGG 240
Qy 181 AGCGACACCTATCCCAAGGCTGCGCGGCGCGGCGGCGGCTGCGCTCAGCCCGG 240

Db 241 TACCTTGGCCCTCTATGCGCATGAGGCGATGGGTGGCGAGATGGCTCTGTCAACC 300
Qy 241 TATCTTGGCCCTCTATGCGCATGAGGCGATGGGTGGCGAGATGGCTCTGTCAACC 300

Db 301 CGCGGCTCTGGGCTAGTTGGGCGCC 327
Qy 301 CGCGGCTCTGGGCGCAATGGGCGCC 327

RESULT 2

ID PCT-US95-10398-113 STANDARD; DNA; UNC; 573 BP.

AC xxxxxx

DT 01-JAN-1900

DE Sequence 113, Application PCT/US95/10398.
CC Sequence 113, Application PCT/US95/10398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 113:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: P10
SQ Sequence 573 BP; 96 A; 185 C; 179 G; 113 T; 0 other;

Query Match 89.9%; Score 285; DB 13; Length 573;
Best Local Similarity 90.8%; Pred. No. 2.94e-183;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGCTACACACCGCCGCCACAG 60
Qy 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGCTACACACCGCCGCCCTCAK 60

Db 61 GACGTCAAGTTCGCGCGGTGTCAGATCGTTGCGAGTTTACCTGTGCGCGCAGG 120
Qy 61 GCGTNNNNNNCCGGTGGCGGTGATCGTTGTCGCTTACCTGTGCGCGCAGG 120


```
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 117:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: IND3
CC Sequence 573 BP; 100 A; 184 C; 174 G; 115 T; 0 other;

Query Match      88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db      1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
Qy      1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTAK 60

Db      61 GAGCTCAAGTTCCGGCGGTGCCAGATCGTTGGTGGAGTTTACCTGTCGCCGCCAGG 120
Qy      61 GSGGTNNNNNNCCGGTGGCGGTGAGATCGTTGGTGGAGTTTACCTGTCGCCGCCAGG 120

Db      121 GCGCCAGGTTGGGTGTGCGCGCGCAGTACGAGCACTTCGAGCGGTGCGTCTGCACCC 300
Qy      241 TATCCTTGGCCCTCTATGCGCAATGAGGGCTGCGGGTGGGGGGTGGCTGCTGTCGCCCC 300

Db      301 CGCGGTCTCGGCTAGTTGGGCCCC 327
Qy      301 CGCGGTCTCGGCTAGTTGGGCCCC 327

RESULT      5
ID      PCT-US95-10398-118 STANDARD; DNA; UNC; 573 BP.
AC      xxxxxx
DT      01-JAN-1900
DE      Sequence 118, Application PC/TUS9510398.
```

```
CC Sequence 118, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PORCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 118:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: IND8
CC Sequence 573 BP; 97 A; 185 C; 177 G; 114 T; 0 other;

Query Match      88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db      1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
Qy      1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTAK 60

Db      61 GAGCTCAAGTTCCGGCGGTGCCAGATCGTTGGTGGAGTTTACCTGTCGCCGCCAGG 120
Qy      61 GSGGTNNNNNNCCGGTGGCGGTGAGATCGTTGGTGGAGTTTACCTGTCGCCGCCAGG 120

Db      121 GCGCCAGGTTGGGTGTGCGCGCGCAGTACGAGCACTTCGAGCGGTGCGTCTGCACCC 180
```

```
Qy 121 GGGCCAGGNGGGTGTGCGCGGACTAGGAAGACTTCGAGCGGTCAACAACCTCGTGGC 180
Db 181 AGCGACAACTATCCCAAGGCTCCGCGGCGGAGGTAGGGCTTGGGCTCAGCCGGG 240
Qy 181 AGCGACAGCTATCCCAAGGCTCGYCGGCGGAGGAGGCTCTGGGCTCAGCCGGG 240
Db 241 CACCTTGGCCCTCTATGCGAATAGGGCTTGGGCTGGGAGAGTGGCTCTCTCACCC 300
Qy 241 TATCTTGGCCCTCTATGCGAATAGGGCTTGGGCTGGGCGGNTGGCTCTCTGCCCC 300
Db 301 CGCGGCTCTGGGCTAGTGGGCGCC 327
Qy 301 CGCGGCTCTGGGCGCAATTGGGCGCC 327
```

RESULT 6
ID PCT-US95-10398-115 STANDARD; DNA; UNC; 573 BP.

```
AC xxxxx
DT 01-JAN-1900
DE Sequence 115, Application PC/TUS9510398.
CC Sequence 115, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/250/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2076-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 115:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
```

```
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: T10
SQ Sequence 573 BP; 99 A; 181 C; 176 G; 117 T; 0 other;

Query Match      88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATCAGCAGCAATCTTAACCTCAAAAGAAAACCAAAAGTAACCAACGCGCGCCACAG 60
Qy 1 ATCAGCAGCAATCTTAACCTCAAAAGAAAACCAAAAGTAACCAACGCGCGCCCTCAK 60

Db 61 GAGGTCAAGTTCGCGGGGGTGGTCAGATCGTTGGTGGAGTTTACTGTTCCGCGCGAGG 120
Qy 61 G:SGTNNNNNNNCGGCTGGCGGTGCGATCGTTGGTGGAGTTTACTGTTCCGCGCGAGG 120

Db 121 GCGCCAGGTGGGTGTCGCGCGGACTAGGAAGACTTCCGAGCGGTCCGAACCTCGTGA 180
Qy 121 GCGCCAGGNGGGTGTGCGCGGACTAGGAAGACTTCCGAGCGGTCAACAACCTCGTGC 180

Db 181 AGCGACAGCTATCCCAAGGCTCCGAGCGGAGGAGGCTTGGGCTCAGCCCGGG 240
Qy 181 AGCGACAGCTATCCCAAGGCTCGYCGGCGGAGGAGGCTTGGGCTCAGCCCGGG 240

Db 241 TACCTTGGCCCTCTATGCGAATAGGGCTGGGCTGGGAGAGTGGCTCTCTGTCAACC 300
Qy 241 TATCTTGGCCCTCTATGCGAATAGGGCTGGGCTGGGAGGNTGGCTCTCTGTCCCCC 300

Db 301 CGTGGCTCCGCGGCTAGTTGGGCGCC 327
Qy 301 CGCGGCTCTCGGCGCAATTGGGCGCC 327
```

RESULT 7
ID PCT-US95-10398-111 STANDARD; DNA; UNC; 573 BP.

```
AC xxxxx
DT 01-JAN-1900
DE Sequence 111, Application PC/TUS9510398.
CC Sequence 111, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
```

```
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 111:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: D1
SQ Sequence 573 BP; 96 A; 184 C; 177 G; 116 T; 0 other;

Query Match      88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db      1  ATGAGCAGCAATCTTAAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCACAG 60
        |||||||
Qy      1  ATGAGCAGCAATCTTAAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db      61  GAGCTCAAGTTCCGGCGGTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120
        1:!! |||||
Qy      61  GGGTNNNNNNCCGGGTGGCGGTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120

Db      121  GCGCCAGGTTGGGTGCGCGGCACTAGGAGACTTCCGAGCGGTCCGCAACCTCGTGGC 180
        |||||||
Qy      121  GCGCCAGGNNGGGTGCGCGGCACTAGGAGACTTCCGAGCGGTCCGCAACCTCGTGGC 180

Db      181  AGCGGACACCTATCCCAAGCTCGCGGCACTAGGAGACTTCCGAGCGGTCCGCAACCTCGCGG 240
        |||||||
Qy      181  AGCGGACAGCTATCCCAAGGCTCGYCGYCGGCGGCGAGGCTCTGGGCTCAGCGCGG 240

Db      241  TACCCTTGGCCCTCTATGGCAACGAGGCTTGGGTGGGCAAGATGGCTCTGTACCC 300
        |||||||
Qy      241  TATCTTGGCCCTCTATGGCAATGAGGCTGCGGGTGGCGGNTGGCTCTGTCCCCC 300

Db      301  CGCGGCTCCCGGCTAGTTGGGGCCCC 327
        |||||||
Qy      301  CGCGGCTCGGCGCAATTTGGGGCCCC 327

RESULT      8
ID PCT-US95-10398-112 STANDARD; DNA; UNC; 573 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 112, Application PC/TUS9510398.
CC Sequence 112, Application PC/TUS9510398
```

```
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 112:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: US6
SQ Sequence 573 BP; 97 A; 183 C; 178 G; 115 T; 0 other;

Query Match      88.0%; Score 279; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 6.39e-179;
Matches 294; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db      1  ATGAGCAGCAATCTTAAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCACAG 60
        |||||||
Qy      1  ATGAGCAGCAATCTTAAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db      61  GAGCTCAAGTTCCGGCGGTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120
        1:!! |||||
Qy      61  GGGTNNNNNNCCGGGTGGCGGTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120

Db      121  GCGCCAGGTTGGGTGCGCGGCACTAGGAGACTTCCGAGCGGTCCGCAACCTCGTGGC 180
        |||||||
```


CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 124:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: T3
SQ Sequence 573 BP; 98 A; 181 C; 177 G; 117 T; 0 other;

Query Match 88.0%; Score 279; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 6.39e-179;
Matches 294; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCGCCACAG 60
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
Db 61 GACGTTAAGTTCGGGGGTGTCAGATCGTTGGTGAGTTTACCTGTTCGGCGCAGG 120
Qy 61 GSGTNNNNNNCCGGGTGCGGTGAGATCGTTGGTGAGTTTACCTGTTCGGCGCAGG 120
Db 121 GCGCCAGGTTGGGTGTCGCGCGACTAGGAAGACTTCGAGCGGTGCGAAGCTCGTGGA 180
Qy 121 GCGCCAGGNNGGGTGTCGCGCGACTAGGAAGACTTCGAGCGGTGCGAAGCTCGTGGA 180
Db 181 AGCGGACAACTATCCCAAGGCTCGCGCGGAGGTAGGGCTGGGCTAGCGCGGG 240
Qy 181 AGCGGACAGCTATCCCAAGGCTCGCGCGGAGGTAGGGCTGGGCTAGCGCGGG 240
Db 241 TACCTTGGCCCTCTATGGGACAGGCGATGGGTGGGAGGATGGCTCTGTACCC 300
Qy 241 TATCTTGGCCCTCTATGGGATGAGGCTGGGCTGGGCGGNTGGCTCTGTCCCCC 300
Db 301 CGCGGCTCGCGGCTTAATTTGGGGCCCC 327
Qy 301 CGCGGCTCGCGGCAATTTGGGGCCCC 327

RESULT 11
ID PCT-US95-10398-120 STANDARD; DNA; UNC; 573 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 120, Application PC/TUS9510398.
CC Sequence 120, Application PC/TUS9510398
CC GENERAL INFORMATION:

CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 120:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: HK3
SQ Sequence 573 BP; 104 A; 183 C; 172 G; 114 T; 0 other;

Query Match 88.0%; Score 279; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 6.39e-179;
Matches 294; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCGCCACAG 60
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
Db 61 GACGTTAAGTTCGGGGGTGTCAGATCGTTGGTGAGTTTACCTGTTCGGCGCAGG 120
Qy 61 GSGTNNNNNNCCGGGTGCGGTGAGATCGTTGGTGAGTTTACCTGTTCGGCGCAGG 120
Db 121 GCGCCAGGTTGGGTGTCGCGCGCAGGAGACTTCAGAGCGGTGCGAAGCTCGTGGA 180
Qy 121 GCGCCAGGNNGGGTGTCGCGCGCAGTAGGAAGACTTCGAGCGGTGCGAAGCTCGTGGA 180


```
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 121:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: HK5
SQ Sequence 573 BP; 103 A; 183 C; 173 G; 114 T; 0 other;

Query Match      88.0%; Score 279; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 6.39e-179;
Matches 294; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGCTCAK 60

Db 61 GAGGTCAAGTTCCCGGGGGTGTGATGCTGTGGTGGAGTTTACCTGTTCGGCGCGCAGG 120
Qy 61 GGGTNNNNNNCCGGGTGGCGGTGATGCTGTGGTGGAGTTTACCTGTTCGGCGCGCAGG 120

Db 121 GCGCCAGAGTTGGTGTGCGCGGACGAGAGACTTCCGAGCGGTGCGAAGCTCTGCGA 180
Qy 121 GCGCCAGAGTTGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTGCGAAGCTCTGCGC 180

Db 181 AGCGCACACCTATCCCAAGGCTCGCGACCGAGCGGACCTGGGCTCAGCCCGGG 240
Qy 181 AGCGCACAGCTATCCCAAGGCTCGCGGAGGCGGAGGCTGCGGCTCAGCCCGGG 240

Db 241 TATCCTTGGCCCTCTATGGCAATGAGGCTGGGGTGGGCGAGATGGCTCTCTCACCC 300
Qy 241 TATCCTTGGCCCTCTATGGCAATGAGGCTGGGGTGGGCGGAGTGGCTCTCTCCGCC 300

Db 301 CATGGCTCTCGGCTAGTTGGGCCCC 327
Qy 301 CGCGGCTCTCGGCCCAATTGGGCCCC 327

RESULT 14
ID PCT-US95-10398-106 STANDARD; DNA; UNC; 573 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 106, Application PC/TUS9510398.
CC Sequence 106, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R. H. AND
```

```
CC APPLICANT: PURCELL, R. H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 106:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: SW1
SQ Sequence 573 BP; 91 A; 182 C; 177 G; 117 T; 0 other;

Query Match      87.7%; Score 278; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 3.37e-178;
Matches 293; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGCTCAK 60

Db 61 GAGGTCAAGTTCCCGGGTGGCGGTGATGCTGTGGTGGAGTTTACTTGTTCGGCGCGCAGG 120
Qy 61 GGGTNNNNNNCCGGGTGGCGGTGATGCTGTGGTGGAGTTTACTTGTTCGGCGCGCAGG 120

Db 121 GCGCCAGAGTTGGTGTGCGCGGACGAGAGACTTCCGAGCGGTGCGAAGCTCTGCGA 180
Qy 121 GCGCCAGAGTTGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTGCGAAGCTCTGCGC 180
```

WQ59E4

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MParch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Feb 23 20:52:18 1998; MasPar time 208.41 Seconds

Tabular output not generated.
774.811 Million cell updates/sec

Title: >US-08-836-075A-1
Description: (1-327) from US08836075A.seq
Perfect Score: 317
N.A. Sequence: 1 ATGACGACGATCTGAACCC.....CTCGGCCCAATTGGGGCCCC 327
Comp: TACTCGTGCTTAGGATTGG.....GAGCGGGTTAACCCGGGG

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 665703 seqs, 246912890 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: EST-A
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
EST-B
99:EST99 100:EST100 101:EST101 102:EST102 103:EST103
104:EST104 105:EST105 106:EST106 107:EST107 108:EST108
109:EST109 110:EST110 111:EST111 112:EST112 113:EST113
114:EST114 115:EST115 116:EST116 117:EST117 118:EST118
119:EST119 120:EST120 121:EST121 122:EST122 123:EST123
124:EST124 125:EST125 126:EST126 127:EST127 128:EST128
129:EST129 130:EST130 131:EST131 132:EST132 133:EST133

134:EST134 135:EST135 136:EST136 137:EST137 138:EST138
139:EST139 140:EST140 141:EST141 142:EST142 143:EST143
144:EST144 145:EST145 146:EST146 147:EST147 148:EST148
149:EST149 150:EST150 151:EST151 152:EST152 153:EST153
154:EST154 155:EST155 156:EST156 157:EST157 158:EST158
159:EST159 160:EST160 161:EST161 162:EST162 163:EST163
164:EST164 165:EST165 166:EST166 167:EST167 168:EST168
169:EST169 170:EST170 171:EST171 172:EST172 173:EST173
174:EST174 175:EST175 176:EST176 177:EST177 178:EST178
179:EST179 180:EST180 181:EST181 182:EST182 183:EST183
184:EST184 185:EST185 186:EST186 187:EST187 188:EST188
189:EST189 190:EST190 191:EST191 192:EST192 193:EST193
194:EST194 195:EST195 196:EST196

Statistics: Mean 9.769; Variance 2.141; scale 4.562

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description	Pred. No.
C	1	23	7.3	423 13	T96691	ye52h03.sl Homo sapie	1.30e-05
	2	22	6.9	301 36	R77276	yi75d01.sl Homo sapie	1.86e-04
	3	22	6.9	426 92	H79148	yu27g08.sl Homo sapie	1.86e-04
C	4	22	6.9	491 54	R94705	yq43d01.sl Homo sapie	1.86e-04
	5	21	6.6	215 28	R49453	yq68e02.sl Homo sapie	2.43e-03
C	6	21	6.6	217 85	N56240	J9576F Homo sapiens c	2.43e-03
C	7	21	6.6	227 55	R97434	yq53h06.sl Homo sapie	2.43e-03
C	8	21	6.6	304 14	R07673	ye98c07.rl Homo sapie	2.43e-03
	9	21	6.6	332 140	AA038712	mi92c09.rl Soares mou	2.43e-03
C	10	21	6.6	404 147	AA119984	mn34a10.rl Beddington	2.43e-03
C	11	21	6.6	419 92	H81577	yu73b06.sl Homo sapie	2.43e-03
	12	21	6.6	459 37	R82259	yj17g01.rl Homo sapie	2.43e-03
	13	21	6.6	468 183	N53148	yv55g10.sl Soares fet	2.43e-03
	14	20	6.3	210 35	R76212	yi71d11.sl Homo sapie	2.89e-02
C	15	20	6.3	276 68	H40395	yp60e05.rl Homo sapie	2.89e-02
	16	20	6.3	305 59	T31232	EST28924 Homo sapiens	2.89e-02
	17	20	6.3	370 95	ATTS3506	A. thaliana transcrib	2.89e-02
	18	20	6.3	374 178	AA192539	zp91g05.rl Stratagene	2.89e-02
	19	20	6.3	379 29	R33992	yj70c03.sl Homo sapie	2.89e-02
C	20	20	6.3	399 39	H13389	yj72b11.rl Homo sapie	2.89e-02
	21	20	6.3	402 16	T09138	EST07031 Homo sapiens	2.89e-02
C	22	20	6.3	415 45	H17241	ym42f08.rl Homo sapie	2.89e-02
	23	20	6.3	418 2	T59290	yc02c12.rl Homo sapie	2.89e-02
	24	20	6.3	437 47	H50371	yo29b01.rl Homo sapie	2.89e-02
C	25	20	6.3	437 45	H17422	ym40e09.rl Homo sapie	2.89e-02
	26	20	6.3	442 26	R42917	yq14b10.sl Homo sapie	2.89e-02
C	27	20	6.3	443 38	H09205	y198d07.rl Homo sapie	2.89e-02
	28	20	6.3	445 92	H79969	yv79f12.rl Homo sapie	2.89e-02
C	29	20	6.3	474 56	H19576	yn59c01.rl Homo sapie	2.89e-02
	30	20	6.3	478 51	R85141	yo41d07.rl Homo sapie	2.89e-02
C	31	20	6.3	482 121	N93628	zb70e11.sl Soares fet	2.89e-02
	32	20	6.3	514 138	AA028338	mi20f11.rl Soares mou	2.89e-02
	33	20	6.3	514 46	H46973	yo19g12.rl Homo sapie	2.89e-02
	34	19	6.0	81 170	AA155866	zo70c12.rl Stratagene	3.10e-01
	35	19	6.0	331 93	H84640	ys67e06.sl Homo sapie	3.10e-01
	36	19	6.0	338 175	AA181878	zp66a02.sl Stratagene	3.10e-01
	37	19	6.0	345 180	AA022461	ze70a03.sl Soares fet	3.10e-01
	38	19	6.0	391 109	W53431	md44d04.rl Soares mou	3.10e-01
	39	19	6.0	422 158	W81106	zd82b02.sl Soares fet	3.10e-01

40 19 6.0 431 186 W84502 zh53a02.r1 Soares fet 3.10e-01
 c 41 19 6.0 -442 92 H79035 yul-q01.s1 Homo sapie 3.10e-01
 42 19 6.0 462 125 W12927 mb20g11.r1 Soares mou 3.10e-01
 43 19 6.0 470 43 R21500 yq06g10.r1 Homo sapie 3.10e-01
 44 19 6.0 493 160 W87973 mf65a07.r1 Soares mou 3.10e-01
 c 45 19 6.0 512 169 AA015386 mh15d07.r1 Soares mou 3.10e-01

ALIGNMENTS

RESULT 1
 LOCUS T96691 423 bp mRNA EST 27-MAR-1995
 DEFINITION yes2h03.s1 Homo sapiens cDNA clone 121397 3'.
 ACCESSION T96691
 NID q735315
 KEYWORDS EST.
 SOURCE human clone=121397 library=Soares fetal liver spleen INFLS

vector=pT7T3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=-21ml3 Reitel=Pac I Reite2=Eco RI Liver and spleen from a 20 week-post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGATTAATAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaudo.

ORGANISM

Homo sapiens
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 423).

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

COMMENT

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

High quality sequence stops: 323

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source

1..423

/organism="Homo sapiens"

/clone="121397"

BASE COUNT 85 a 110 c 131 g 90 t 7 others

ORIGIN

Query Match 7.3%; Score 23; DB 13; Length 423;

Best Local Similarity 70.5%; Pred. No. 1.30e-05;

Matches 31; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 165 cgggctgagaccctctccctgatcgagccaggttgagg 208

||||||| || | ||||| : | : | |||||

Cp 238 CGGGTGAGCCAGGACCTGCGCTGCGCGCGCTTGGG 195

RESULT 2
 LOCUS R77276 301 bp mRNA EST 06-JUN-1995
 DEFINITION y175d01.s1 Homo sapiens cDNA clone 145057 3'.
 ACCESSION R77276
 NID g851908
 KEYWORDS EST.
 SOURCE human clone=145057 library=Soares placenta Nb2HP vector=pT7T3D

(Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=Promega -21ml3 Reitel=Not I Rsite2=Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGATTCGGCGCGAGGAATTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaudo.

Homo sapiens
 Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.

ORGANISM

1 (bases 1 to 301)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

REFERENCE

The WashU-Merck EST Project

AUTHORS

Unpublished (1995)
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

High quality sequence stops: 276

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source

1..301

/organism="Homo sapiens"

/clone="145057"

BASE COUNT 65 a 56 c 127 g 53 t

ORIGIN

Query Match 6.9%; Score 22; DB 36; Length 301;

Best Local Similarity 84.4%; Pred. No. 1.86e-04;

Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 112 gggccggggcgagccaggtctcagccagg 143

||||| ||||| || ||||| |||||

Qy 209 GYCGGAGGCGAGTCTCGGCTAGCGCGG 240

RESULT 3

LOCUS H79148 426 bp mRNA EST 09-NOV-1995

DEFINITION yu27q08.s1 Homo sapiens cDNA clone 235070 3'.

ACCESSION H79148

NID g1057237

KEYWORDS EST.

FEATURES

```
source
1..227
/organism="Homo sapiens"
/cclone="199547"
```

mrna	<1..>227	
BASE COUNT	50 a	57 c
ORIGIN	76 g	42 t
		2 others

Query Match	6.6%	Score 21	DB 55	Length 227
Best Local Similarity	68.2%	Pred. No. 2.43e-03		
Matches 30	Conservative 2	Mismatches 12	Indels 0	Gaps 0
DB 152	cgggctgagaccgcgtctctgccctgatgcagccagattnggg	195		
		:		
DB 238	cgggctgagccagcagaccttgccttcgcrccgacgctttgggg	195		
		:		

8	RESULT	R07673	304 bp	mRNA	EST	05-APR-1995
	LOCUS	ye98c07.r1	Homo sapiens	cDNA clone	125772	5' similar to contains
	DEFINITION		Alu repetitive element.			
	ACCESSION	R07673				
	NTID	q759596				
	KEYWORDS	EST.				
	SOURCE	human clone=125772 library=Soares fetal liver spleen	INFLS			
		vector=pT73D (Pharmacia) with a modified polylinker	host=DHI10B			
		(ampicillin resistant) primer=M13Rp1	Rsite1=Pac I	Rsite2=Eco RI		
		Liver and spleen from a 20 week-post conception male fetus.	1st			
		strand cDNA was primed with a Pac I - oligo (dT) primer [5'				
		AACCTGGACAGATAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'],	double-stranded			
		cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac				
		I and cloned into the Pac I and Eco RI sites of the modified pT73				
		vector. Library went through one round of normalization. Library				
		constructed by Bento Soares and M.Fatima Bonaldo.				

EASE COUNT		ORIGIN	
80 a	78 c	53 g	91 t
		2 others	

Query Match 6.6%; Score 21; DB 14; Length 304;
Best Local Similarity 85.7%; Pred. No. 2.43e-03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

RESULT	9
LOCUS	AA038712 332 bp mRNA EST 28-AUG-1996
DEFINITION	mi92c09.r1 Soares mouse p3NNF19.5 Mus musculus cDNA clone 74064 5' similar to gb:K53476 Mouse mRNA for non-histone chromosomal protein HMG-14 (MOUSE);.

ACCESSION	AA038712
NID	g1514121
KEYWORDS	EST.

SOURCE house mouse.
ORGANISM *Mus musculus*
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.

```

FEATURES
source
Location/Qualifiers
1..332
/organism="Mus musculus"
/note="Vector: pT7T3D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCTTACCAATCTGAAGTCGACGCCGCGCATTTTTTTTTTTTTTTT 3')],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Ronaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)".
/clone="747064"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
<1..>332
83 a 36 c 92 g 101 t
mrna
BASE COUNT
ORIGIN

```

Query Match 6.6%; Score 21; DB 140; Length 332;
 Best Local Similarity 92.0%; Pred. No. 2.43e-03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 205 tcaaaagaaaacacatcatcacc 229
 |||||
 Cp 21 TCAAGAAAACCAACGTAACACC 45

RESULT 10
 LOCUS AA119984 404 bp mRNA EST 19-NOV-1996
 DEFINITION mn34a10.r1 Beddington mouse embryonic region Mus musculus cDNA
 clone 539802 5' similar to TR:G1236085 G1236085 LISCH7 ;
 ACCESSION AA119984
 NID q1677684
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Eukaryotes; Metazoa; Chordata;
 Vertebrata; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.

REFERENCE 1 (bases 1 to 404)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 TITLE The HashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT

Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:326738

Putative full length read
 vector to vector length is 425
 Seq primer: -40ml3 ET
 High quality sequence stop: 362.

FEATURES
 source
 1..404
 /organism="Mus musculus"
 /strain="C57BL6 x DBA"
 /note="Vector: pCMV-SPORT; Site 1: SalI; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dT. Gastrulating
 embryos were collected at 7.5dpc from C57BL6 x DBA
 matings, excluding embryos that had developed head folds
 and all extraembryonic tissues. Average insert size: 1.3
 kb (range: 0.5 - 3.0 kb). Referenced in Development 121,
 2479-2489 (1995)"
 /clone="539802"
 /clone_lib="Beddington mouse embryonic region"
 /sex="pooled"
 /dev_stage="7.5dpc"
 /lab_host="DH12S"
 <1..>404

mRNA
 BASE COUNT 107 a 94 c 109 g 94 t
 ORIGIN

Query Match 6.6%; Score 21; DB 147; Length 404;
 Best Local Similarity 66.7%; Pred. No. 2.43e-03;
 Matches 39; Conservative 1; Mismatches 18; Indels 0; Gaps 0;

Db 44 ggtctaggatccagtcgctgacccagatccgctcagcgatccacgacatccctcggg 100
 |||||
 Cp 252 GGGCAAGGATACCGGGCTGAGCCAGGACCTGCCCTCGGRCGCCGCCCTGGG 196

RESULT 11
 LOCUS H81577 419 bp mRNA EST 11-DEC-1995
 DEFINITION yu73b06.s1 Homo sapiens cDNA clone 239411 3'.
 ACCESSION H81577
 NID g1114654
 KEYWORDS EST.
 SOURCE human clone=239411 primer=Promega -2ml3 library=Weizmann Olfactory
 Epithelium vector=pBluescript SK- host=SOLR cells (kanamycin
 resistant) Reitel=EcoRI Reitel=XhoI From 35 year old female. The
 cDNA was oligo (dT) primed with an XhoI restriction enzyme
 recognition site and an 18 base poly dT sequence. For the 5' end,
 the synthesized cDNA termini were treated with T4 DNA polymerase
 and EcoRI adaptors were ligated to the blunt ends.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcophylli; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
 Euthera; Archonta; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 419)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
 TITLE WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 294
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
 source
 1..419
 /organism="Homo sapiens"
 /clone="239411"
 <1..>419
 BASE COUNT 121 a 70 c 62 g 164 t 2 others
 ORIGIN

Query Match 6.6%; Score 21; DB 92; Length 419;
 Best Local Similarity 61.5%; Pred. No. 2.43e-03;
 Matches 40; Conservative 2; Mismatches 22; Indels 1; Gaps 1;

Db 46 ggaatgcaaacctcatgagcgaggtttgtttgtttatgtattgtttacaccaggatt 105
 |||||
 Cp 74 GGNNNNNACWCMTCAGGGCGCGGT--GGTGTACGTGGTGGTTCTTTGAGGTTT 16

Db 106 agaatt 110

_____ (TM)

Database:

RESULT 1

ALIGNMENTS

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description	Pred. No.
c	1	22	6.9	226	98	AA282971	zt14a08.s1 Soares Nbh	9.16e-05
c	2	22	6.9	326	42	HSPD03892	H.eapiens EST sequenc	9.16e-05
c	3	22	6.9	353	65	AA395721	28123 Lambda-PRL2 Ara	9.16e-05
c	4	22	6.9	458	116	AA454719	zx7b02.s1 Soares ova	9.16e-05
c	5	21	6.6	309	50	AA231797	Zt630.f1 cDNA from ric	1.20e-03
c	6	21	6.6	369	50	AA004946	zh87e10.s1 Soares fet	1.20e-03
c	7	21	6.6	389	9	AA073040	mm78b02.r1 Stratagene	1.20e-03
c	8	21	6.6	421	89	AA347253	zv54e08.s1 Soares tes	1.20e-03
c	9	21	6.6	427	94	AA126152	zt185h07.r1 Stratagene	1.20e-03
c	10	21	6.6	450	81	AA410657	zt29g04.r1 Soares ova	1.20e-03
c	11	21	6.6	473	79	AA402119	zv55d04.r1 Soares ova	1.20e-03
c	12	20	6.3	216	61	AA031467	zk16c05.r1 Soares pre	1.44e-02
c	13	20	6.3	259	65	AA396794	vb28h10.r1 Soares mou	1.44e-02
c	14	20	6.3	291	9	AA076835	7B03F10 Chromosome 7	1.44e-02
c	15	20	6.3	312	66	AA403384	wf79f11.r1 Soares mou	1.44e-02
c	16	20	6.3	331	67	AA409505	ESR01244 Mouse 7.5 dp	1.44e-02
c	17	20	6.3	358	6	AA221093	mv75g01.r1 Soares mou	1.44e-02
c	18	20	6.3	378	85	AA428428	zvf4c12.r1 Soares ova	1.44e-02
c	19	20	6.3	382	95	AA147251	zl50d06.r1 Soares pre	1.44e-02
c	20	20	6.3	415	24	AA290473	vb37a07.r1 Soares mou	1.44e-02
c	21	20	6.3	433	16	AA080174	mm34c07.r1 Stratagene	1.44e-02
c	22	20	6.3	435	97	AA278629	zs78h01.r1 Soares Nbh	1.44e-02
c	23	20	6.3	453	97	AA215509	zr97b06.r1 Soares Nbh	1.44e-02
c	24	20	6.3	473	95	AA147482	zl50a05.r1 Soares pre	1.44e-02
c	25	20	6.3	500	47	AA271194	vb79c09.r1 Soares mou	1.44e-02
c	26	19	6.0	133	10	AA23407	zr29e02.r1 Stratagene	1.56e-01
c	27	19	6.0	161	44	AA261679	mz62d07.r1 Soares mou	1.56e-01
c	28	19	6.0	171	114	AA449928	zx37h01.r1 Soares tot	1.56e-01
c	29	19	6.0	253	105	HUM168004B	Human fetal brain cDN	1.56e-01
c	30	19	6.0	255	40	AA471715	ESR41715 Endometrial	1.56e-01
c	31	19	6.0	271	76	AA353160	ESR61251 Activated T-	1.56e-01
c	32	19	6.0	284	41	AA340074	ESY45224 Fetal skin H	1.56e-01
c	33	19	6.0	306	9	AA078095	7P01E12 Chromosome 7	1.56e-01
c	34	19	6.0	340	38	AA330985	ESR34774 Embryo, 6 we	1.56e-01
c	35	19	6.0	341	35	AA321076	ESR23511 Adipose tiss	1.56e-01
c	36	19	6.0	438	116	AA4455957	sl16g06.s1 Soares NHH	1.56e-01
c	37	19	6.0	355	98	AA281664	zt06g04.r1 Soares Nbh	1.56e-01
c	38	19	6.0	362	110	RICC1690A	Rice cDNA, partial se	1.56e-01
c	39	19	6.0	380	75	AA349543	ESY56223 Infant brain	1.56e-01
c	40	19	6.0	390	13	AA243729	zr67b12.s1 Soares Nhh	1.56e-01
c	41	19	6.0	410	79	AA402016	zx53b11.s1 Soares ova	1.56e-01
c	42	19	6.0	440	104	HUM103G04B	Human fetal brain cDN	1.56e-01
c	43	19	6.0	448	66	AA403509	mz55e10.r1 Barstead m	1.56e-01
c	44	19	6.0	483	110	RICC2301A	Rice cDNA, partial se	1.56e-01
c	45	19	6.0	496	110	RICR0152A	Rice cDNA, partial se	1.56e-01

LOCUS AA282971 226 bp mRNA EST 15-MAY-1997
 DEFINITION zt14a08.s1 Soares NbHTGC Homo sapiens cDNA clone 713078 3'.
 ACCESSION AA282971
 NID g1925885
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
 Homo.
 REFERENCE 1 (bases 1 to 226)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,K., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WASHU-Merck EST Project 1997
 Unpublished (1997)
 TITLE
 JOURNAL
 COMMENT Contact: Wilson RK
 WASHU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -41m13 fwd. ET from Amersham
 High quality sequence stop: 220.

FEATURES
 source
 1..226
 /organism="Homo sapiens"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from germinal B-cells (flow-sorted from
 tonsils) provided by Dr. Louis Staudt of the NCI, and was
 then primed with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAAGTCGAGCGCGCTCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of formalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 /clone="713078"
 /tissue_type="Germinal B-cell"
 /lab_host="DH10B"
 complement(1..>226)
 mRNA 42 a 56 c 70 g 58 t
 BASE COUNT
 ORIGIN

Query Match 6.9%; Score 22; DB 98; Length 226;
 Best Local Similarity 70.6%; Pred. No. 9.16e-05;
 Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 Db 126 tccctgtgtccatctctgtgtaggacgtggagtggtctgtggtctgt 176
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Qy 243 TCCTGGCCCTCTATGCGCATGAGGGCTGCGGGTGGGGGTTGCTCT 293
 RESULT 2
 LOCUS HSPD03892 326 bp DNA EST 21-APR-1997
 DEFINITION H.sapiens EST sequence (007-X3-17) from skeletal muscle.
 ACCESSION F19292

NID g1135707
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
 Homo.
 REFERENCE 1 (bases 1 to 326)
 AUTHORS Lanfranchi,G., Muraro,T., Caldara,F., Pacchioni,B., Pallavicini,A., Pandolfo,D., Toppo,S., Trevisan,S., Scarsio,S. and Valle,G.
 Identification of 4,370 expressed sequence tags (ESTs) from a
 3'-end specific cDNA library of human skeletal muscle by DNA
 sequencing and filter hybridisation
 Genome Res. 1, 35-42 (1996)
 REFERENCE 2 (bases 1 to 326)
 AUTHORS Valle,G.
 TITLE Direct Submission
 JOURNAL Submitted (19-DEC-1995) CRIBI Biotechnology Centre, University of
 Padua, Via Trieste 75, 35121 Padua, Italy
 COMMENT ABI Chromatograms and other information are available on WWW at
 http://eos.bio.unipd.it
 Vector: pcDNAII (Invitrogen)
 V_type: Plasmid
 RE_1: EcoRI
 RE_2: NotI
 PRIMER: PCR
 DESC: The cDNA was constructed using a biotin-NotI-oligo(dT)
 primer,
 sonicated and size selected (450-550 BP). The biotinylated 3'-ends
 were affinity purified, directionally cloned and sequenced (5' ->
 3').

FEATURES
 source
 Location/Qualifiers
 1..326
 /organism="Homo sapiens"
 /note="caucasian"
 /tissue_type="pectoral muscle (after mastectomy)"
 /clone_lib="HM1"
 /sex="female"
 BASE COUNT 60 a 104 c 92 g 69 t 1 others
 ORIGIN
 Query Match 6.9%; Score 22; DB 42; Length 326;
 Best Local Similarity 70.5%; Pred. No. 9.16e-05;
 Matches 31; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 Db 128 ccccaatcctgcctgcacaggcaggacggggtctcagcccg 171
 ||||| ||| ||| ||| ||||| ||| ||||| ||| ||||| |||
 Qy 195 CCCCAAGGCTCGYGGYCGGCGGCGGCGGCTCTGGGCTCAGCCCG 238

RESULT 3
 LOCUS AA395721 353 bp mRNA EST 25-APR-1997
 DEFINITION 28123 Lambda-PRL2 Arabidopsis thaliana cDNA clone 40B12XP 3'.
 ACCESSION AA395721
 NID g2048946
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
 Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
 plants; Magnoliophyta; Magnoliopsida; Capparales; Braesicaceae;
 Arabidopsis.
 REFERENCE 1 (bases 1 to 353)
 AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,

McIntosh, J., Ohlrogge, J., Raikhel, N., Somerville, S., Thomasow, M., Retzel, E. and Somerville, C.
Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)
JOURNAL
MEDLINE 95148729

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@ibm.cl.msu.edu
Seq primer: M13 -21 dve primer.

FEATURES	source
Location/Qualifiers	
1..353	
/organism="Arabidopsis thaliana"	
/strain="var columbia"	
/note="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not; lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is PRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."	
/clone="40B12XP"	
/clone_lib="Lambda-PRL2"	
complement(<1..>353)	
84 a	58 c 75 g 122 t 14 others
ORIGIN	
BASE COUNT	
mRNA	

Query Match 6.9%; Score 22; DB 65; Length 353;
Best Local Similarity 68.2%; Pred. No. 9.16e-05;
Matches 30; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db	84	ggcngngttaa	acctgatgc	natgcagg	gccgcgga	agtgtgt	127
QV	94	gTGGAGT	TAACTGTT	CCGCGAG	SGGCCCA	GNNGGTGT	137

RESULT	4	LOCUS	AA454719	458 bp	mRNA	EST	06-JUN-1997
EFINITION		zz17b02.s1	Soares ovary tumor	NbHOT	Homo sapiens	cDNA clone	809739

3385A710

ACCESSION	AA434719
NID	g2177495

KEYWORDS EST.
SOURCE human

ORGANISM Homo sapiens.

Eukaryota
Vertebrates

Homo.

REFERENCE
HILLIER, I. (base)

Kucaba, T.

Moore, B.,
White, Y.,

TITLE	WashU-Mer	Health Lib
TOURNAY		

JOURNAL OF PUBLICATIONS

COMMENT:

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -41m3 fwd. ET from AmerSham
High quality sequence stop: 406.

```

FEATURES
source
Location/Qualifiers
i..458
/organism="Homo sapiens"
/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTACACATCTCAAGTCGACGCGCGCGGTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). library constructed by Bento Soares and
M.Fatima Bonaldo."
/clone="809739"
/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
complement (<1..>458)
/db_xref="CDB:6039448"
93 a 126 c 145 g 94 t

```

Query Match 6.9%; Score 22; DB 116; Length 458;
Best Local Similarity 70.5%; Pred. No. 9,16e-05;
Matches 31; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 165 cgggctgagaccgcgtcctgccctgatcgaggccaggattgggg 208
 ||||| || | ||||| : | : || |||||
 Cp 238 CGGGGTGAGCCCGAGGACTCGCCTCGGCGCGCGAGCCTTGGGG 195

RESULT	5	AA231797	309 bp	mRNA	EST
LOCUS					

DEFINITION RZ630.F cDNA from rice *Oryza sativa* cDNA clone RZ630.

ACCESSION AA231797
NID Q1854169

KEYWORDS EST.

SOURCE	PRICE.
ORGANISM	Oryza sat

Eukaryota
Chlorobut

Liliopsida

REFERENCE 1 (bases
AUTHORS VanDevnzer

Cartinhov

TITLE	Anchor Point	Unpublished
JOURNAL		

COMMENT

Contact: _____
Dept Plan _____

Cornell U.

Ithaca, I

Tel: 607 255 0420
 Fax: 607 255 6683
 Email: arm4@cornell.edu
 cDNA from rice (Oryza sativa); forward sequence of RFLP probe
 RZ630. Sequence determined by Nicola M. Ayres.

FEATURES

source
 1..309 Location/Qualifiers
 /organism="Oryza sativa"
 /cultivar="IR36"
 /note="Vector: Lambda ZAP II/pBluescript; Site_1: EcoRI; A
 Lambda ZAP II cDNA library was constructed from mRNA
 extracted from etiolated leaf tissue of the rice cultivar
 'IR36' and converted to pBluescript (amp resistant) as
 described in Cause et al. (1994) Genetics 138:1251-1274.
 For insert amplification, use M13 forward and reverse
 primers. Clones from this library are designated with the
 prefix 'RZ'.
 /clone="RZ630"
 /clone_lib="cDNA from rice"
 <1..>309

mRNA

BASE COUNT 73 a 80 c 74 g 82 t
 ORIGIN

Query Match 6.6%; Score 21; DB 8; Length 309;
 Best Local Similarity 62.5%; Pred. No. 1.20e-03;
 Matches 30; Conservative 1; Mismatches 17; Indels 0; Gaps 0;
 Db 211 caacaagacctccagccacccggccacatgacgcgagggggc 258
 ||||| ||| | ||||| ||| | ||| ||||| |||
 Cp 97 CACCAACACTGTCACCGCCACCGGNNNNNNNNACMCMTGAGGCGGC 50

RESULT 6
 LOCUS AA004946 369 bp mRNA EST 07-MAY-1997
 DEFINITION zh87e10.s1 Soares fetal liver spleen INFLS S1 Homo sapiens cDNA
 clone 428298 3'.
 ACCESSION AA004946
 NID g1447733
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
 Homo.
 REFERENCE 1 (bases 1 to 369)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Ellington, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
 Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LIND ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 760 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 215.

FEATURES

source
 1..369 Location/Qualifiers
 /organism="Homo sapiens"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site 2: Eco RI;
 This is a subtracted version of the original Soares fetal
 liver spleen INFLS library. 1st strand cDNA was primed
 with a Pac I - oligo(dT) primer [5'
 AACTGAGCAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."
 /clone="428298"
 /clone_lib="Soares fetal liver spleen INFLS S1"
 /sex="male"
 /dev stage="20 week-post conception fetus"
 /lab host="DH10B (ampicillin resistant)"
 complement (<1..>369)
 /db xref="GDB:1328067"

BASE COUNT 74 a 77 c 120 g 95 t 3 others
 ORIGIN

Query Match 6.6%; Score 21; DB 50; Length 369;
 Best Local Similarity 70.2%; Pred. No. 1.20e-03;
 Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 Db 179 tgggtggagtgagcaggtgccttcgagggccagtcgaggggtgc 225
 ||||| ||| | ||||| ||||| ||||| ||||| |||||
 Qy 93 TGGTGGAGTTTACTTGTTCGCCGCCAGGGGCCGCCAGNGGGGTGTC 139

RESULT 7
 LOCUS AA073040 389 bp mRNA EST 07-FEB-1997
 DEFINITION mm78b02.r1 StrataGene mouse embryonic carcinomara (#937318) Mus
 musculus cDNA clone 534507 5' similar to TR:G1236085 G1236085
 LISCH7 i.
 ACCESSION AA073040
 NID g1594770
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.
 REFERENCE 1 (bases 1 to 389)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT

Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouse@watson.wustl.edu
 This clone is available royalty-free through LIND ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:321443	
Seq primer: -28ml3 rev1 ET from Amersham.	
Location/Qualifiers	
source	1..389

1000 bp. The template was constructed as follows: pVector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. p19 cell line treated with retinoic acid. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'-GAAATTCGACACAG 3' ~3' adaptor sequence: 5'-GAAATTCGACACAG 3'.

/clone lib="Stratagene mouse embryonic carcinomaRA"
 (#937318)"
 /dev stage="RA-treated P19 cell line (ATCC CRL1825)"
 /lab_host="SOLR (kanamycin resistant)"
 <!--389

| ORIGIN | EASE COUNT | mRNA |
|--------|------------|----------|
| 98 a | 105 c | <1..>389 |
| 117 g | | |
| 69 t | | |

Query Match 6.6%; Score 21; DB 9; Length 389;
Best Local Similarity 66.7%; Pred. No. 1.20e-03;
Matches 38; Conservative 1; Mismatches 18; Indels 0; Gaps 0;

Db 204 ggtctaggatccacgtgctgacccagatcccgta,agtccacgatacctcggg 260

Cp 252 GGGCCAAGGATACCCGGGCTGAGCCCAGGACCTGCCCTCGGCCGRCGAGCCTTGGG 196

| RESULT | 8 | LOCUS | AA437253 | 421 bp | mrna | EST | 30-MAY-1997 |
|------------|---|-------|------------|--------|--------|-----|-----------------------------------|
| DEFINITION | | | z554608.s1 | Soares | testis | NHT | Homo sapiens cDNA clone 757478 3' |
| ACCESSION | | | AA437253 | | | | |
| | | | NID | | | | |
| KEYWORDS | | | g2142167 | | | | |
| | | | EST. | | | | |
| SOURCE | | | human. | | | | |

ORGANISM Homo sapiens
mitochondrial eukaryotes; Metazoa; Chordata;
Eukaryotae; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
Homo.

REFERENCE
AUTHORS
1 (bases 1 to 421)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marrs, M., Martin, J.,
Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.

| TITLE | JOURNAL | WashU-Merck EST Project 1997 | Unpublished (1997) |
|-------|---------|------------------------------|--------------------|
| | | | |

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -41m3 fwd. ET from Amersham
High quality sequence stop: 291.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1..421 |

```

/organism="Homo sapiens"
/note="Vector: pT7D3D-pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand
cDNA

```

was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTCACCACTCAAGTGGAGGGCGGCCCAATTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Col5, and was constructed by Bento Soares and M. Fatima Bonaldo.*
/clone="757478"

```
/clone_lib="Soares testis NHT"  
/sex="male"
```

```

/lab_host="DH10B"
complement(<1..>421)
/db_xref="CDB:597809

```

| BASE COUNT | 75 a | 122 c | 151 g | 73 t |
|------------|------|-------|-------|------|
| ORIGIN | | | | |

Query Match 6.6%; Score 21; DB 89; Length 421;
Best Local Similarity 82.8%; Pred. No. 1.20e-03;
Matches 24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 377 gctgggccgagggcacgtcctgggcacag 405

| RESULT | LOCUS | 427 bp | mRNA | EST | 19-MAY-1997 |
|--------|------------|----------|------------|----------------------------|-----------------------------|
| 9 | DEFINITION | AA126152 | z18sh07.r1 | Stratagene colon (#937204) | Homo sapiens cDNA clone |
| | | | 511453 5' | similar to TR:G451854 | G451854 ALPHA-MANNOSIDASE ; |

ACCESSION RA126152
 NID gl685800
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
 Homo.

| | |
|-----------|---|
| REFERENCE | 1 (bases 1 to 427) |
| AUTHORS | Millier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, I., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasik, E., Watson, R., Williamson, A., Wohldmann, P., and Wilson, R. |
| TITLE | WashU-Merck EST Project |
| | Unpublished (1995) |

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2041 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 240.

```

FEATURES
source
Location/Qualifiers
1..427
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:

```

Oligo dT. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'.

/clone="511453"

/clone_lib="Stratagene colon (#937204)"

/lab_host="SOLR cells (kanamycin resistant)"

<1..>427

/db_xref="GDB:3843948"

BASE COUNT 74 a 125 c 147 g 76 t 5 others

ORIGIN

Query Match

Best Local Similarity 6.6%; Score 21; DB 94; Length 427;

Matches 20; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Db 219 ccctcagtggtgctgcccgcgggtggcg 248

|||||: ||| |||||||||

Qy 54 CCCTCAGCGCTNNNNNNCCGGTGGCGG 83

RESULT 10

LOCUS AA410657 450 bp mRNA EST 18-MAY-1997

DEFINITION zt29g04.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 723798 5'.

ACCESSION AA410657

NID g2069762

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;

Homo.

REFERENCE 1 (bases 1 to 450)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-Merck EST Project 1997

Unpublished (1997)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 443.

Location/Qualifiers

1..450

/organism="Homo sapiens"

/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaudo."

/clone_lib="Soares ovary tumor NbHOT"

/clone="723798"

/clone_lib="Soares ovary tumor NbHOT"

/sex="Female"

/tissue_type="ovarian tumor"

/lab_host="DH10B (ampicillin resistant)"

<1..>450

/db_xref="GDB:5935183"

BASE COUNT 86 a 135 c 144 g 83 t 2 others

ORIGIN

Query Match

Best Local Similarity 82.8%; Pred. No. 1.20e-03;

Matches 24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 273 cctgcccaggagctgcctcgcccgccagc 301

|||||: ||| |||||||||

Cp 234 CTGAGCCGAGGACCTGCCCTCGGCGCGC 206

RESULT 11

LOCUS AA402119 473 bp mRNA EST 16-MAY-1997

DEFINITION zn55404.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 741895 5' similar to FR:G397579 G397579 L15 MRNA. ;.

ACCESSION AA402119

NID g2056110

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;

Homo.

REFERENCE 1 (bases 1 to 473)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-Merck EST Project 1997

Unpublished (1997)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 462.

Location/Qualifiers

1..473

/organism="Homo sapiens"

/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaudo."

/clone="741895"

/clone_lib="Soares ovary tumor NbHOT"

Feb 23 20:55

home fuller/feb1098/US-08-836-075A-1.rstb

13

```
/sex="Female"
/tissue type="ovarian tumor"
/lab host="DH10B (ampicillin resistant)"
<1..>473
/db xref="GDB:5942137"
BASE COUNT 88 a 153 c 147 g 85 t
ORIGIN
```

```
Query Match 6.6%; Score 21; DB 79; Length 473;
Best Local Similarity 82.8%; Pred. No. 1.20e-03;
Matches 24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 260 ctgtgccaggagctgcccctggccagc 288
|||||
Cp 234 CTGAGCCGAGGACCTGCCCTGGCGCGC 206
```

```
RESULT 12
LOCUS AA031467 216 bp mRNA EST 09-MAY-1997
DEFINITION zk16c05.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
470696 5'.
ACCESSION AA031467
NID g1501533
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
Homo.

REFERENCE 1 (bases 1 to 216)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 936 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 188.

FEATURES
source
1..216
Location/Qualifiers

/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AACTGGAAGAAATTCGGCGCGCTTTTTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/clone="470696"
/clone lib="Soares pregnant uterus NbHPU"
/sex="Female"

Feb 23 20:55

home fuller/feb1098/US-08-836-075A-1.rstb

14

```
/dev stage="adult"
/lab host="DH10B"
<1..>216
/db xref="GDB:3756698"
BASE COUNT 41 a 92 c 52 g 29 t 2 others
ORIGIN
```

```
Query Match 6.3%; Score 20; DB 61; Length 216;
Best Local Similarity 72.1%; Pred. No. 1.44e-02;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 22 gccggccagtggaagcagcagccagccaccacccccc 64
|||||
Cp 313 GCCGAGACCGCGGGGGGACAGGAGCCANCCGCCACCGCGCA 271
```

```
RESULT 13
LOCUS AA396794 259 bp mRNA EST 25-APR-1997
DEFINITION vb28h10.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
750307 5'.
ACCESSION AA396794
NID g2049801
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
```

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
Murinae; Mus.

REFERENCE 1 (bases 1 to 259)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:459291

Putative full length read
vector to vector length is 260
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES
source
1..259
Location/Qualifiers

/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TCTTACCATCTGAAGTGGAGCGCGGATACCTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M.Fatima Bonaldo."
/clone="750307"

Query Match 6.3%; Score 20; DB 66; Length 312;
Best Local Similarity 72.2%; Pred. No. 1.44e-02;
Matches 26; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Db 135 cccaaggttggccgtctgggtgcaggcctgtgct 170
||||| |!|!| |!!!! |||| |
QY 196 CCAAGGCTCGYCGGYCGAGGCAGGCTCTGGGCT 231

Search completed: Mon Feb 23 20:58:41 1998
Job time : 135 secs.


```
FT mRNA <1..>350
SQ Sequence 350 BP; 71 A; 96 C; 128 G; 55 T; 0 other;

Query Match 6.9%; Score 22; DB 64; Length 350;
Best Local Similarity 70.6%; Pred. No. 1.48e-04;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 110 tccctgggtccatctctggtgagcgtggtggagtgctctggtgctgct 160
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 243 TCCTTGGCCCTCTATGCCAATGAGCGCTCGCGGTGCGCGGTGCTCTCT 293

RESULT 6
ID HSA54719 standard; RNA; EST; 458 BP.
AC AM54719;
NI g2177495
DT 07-JUN-1997 (Rel. 52, Created)
DT 07-JUN-1997 (Rel. 52, Last updated, Version 1)
DE zx77b02.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 809739
DE 3'.
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP 1-458
RA Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,
RA Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J.,
RA Moore B., Schellenberg K., Steptoe M., Tan F., Theising B.,
RA White Y., Wylie T., Waterston R., Wilson R.;
RT *WashU-Merck EST Project 1997*;
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@watson.wustl.edu This clone is available royalty-free through
CC LMLL ; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Seq primer: -41m13 fwd. ET from Amersham High
CC quality sequence stop: 406.
FH Key Location/Qualifiers
FH
FH
FT 1..458
FT /organism="Homo sapiens"
FT /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
FT modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
FT strand cDNA was primed with a Not I - oligo(dT) primer [5'
FT TCTTACCAATCTGAAGTCGAGCGCGCGGTGTTTTTTTTTTT 3'],
FT double-stranded cDNAs size selected, ligated to Eco RI
FT adapters (Pharmacia), digested with Not I and cloned into
FT the Not I and Eco RI sites of a modified pT73 vector
FT (Pharmacia). Library constructed by Bento Soares and
FT M.Fatima Bonaldo."
FT /clones="809739"
FT /clone_lib="Soares ovary tumor NbHOT"
FT /sex="Female"
FT /tissue_type="ovarian tumor"
FT /lab_host="DH10B (ampicillin resistant)"
FT complement(<1...>458)
SQ Sequence 458 BP; 93 A; 126 C; 145 G; 94 T; 0 other;

Query Match 6.9%; Score 22; DB 78; Length 458;
Best Local Similarity 70.5%; Pred. No. 1.48e-04;
Matches 31; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
```

```
Db 165 cgggctgagaccgcctcctgcctgatgcagccaggattgggg 208
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 238 C3GGCTGAGCCGAGCACTGCCTCGGCCGRCGAGCCTTGGGG 195

RESULT 7
LOCUS G25421 294 bp DNA STS 31-MAY-1996
DEFINITION human STS EST405749.
ACCESSION G25421
NID g1347653
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene
collection..
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 294)
AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSs
JOURNAL Unpublished (1995)
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: ACAAAATATTACCAACCCGAAA
Primer B: TTTTATGTACCAATTGCTGAATG
STS size: 150
PCR Profile:
Presoak:
Denaturation: Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3
Derived from dbEST (genbank accession H81577).
FEATURES
source
Location/Qualifiers
1..294
/organism="Homo sapiens"
/map="48.5 cR from top of Chr14 linkage group"
STS
1..150
primer_bind 1..22
primer_bind complement(127..150)
BASE COUNT 90 a 49 c 48 g 107 t
ORIGIN
```


Cp 234 CTGACCCAGGACCTGCCCTCGGRCGC 206

||||| ||||| ||||| || : |

RESULT 10

LOCUS G29737 419 bp DNA STS 04-OCT-1996

DEFINITION human STS SHCC-35139.

ACCESSION G29737

NTD gi593288

KEYWORDS STS sequence; primer; sequence tagged site.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (Bases 1 to 419)

AUTHORS Myers,R.M.

JOURNAL Unpublished (1996)

COMMENT

Contact: Richard M. Myers
Stanford Human Genome Center (SHCC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

Primer A: ACATAATATTACCACCGAAA
Primer B: TTTTATGTACCAATTGCTGAATG

STS size: 150

PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds

Annealing: 62 degrees C for 23 seconds

Polymerization: 72 degrees C for 30 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Taq Polymerase: 0.05 units/uL

Total Vol: 10 uL

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 20 mM

pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from H
-- Washington University/Merck EST sequence.

FEATURES

source Location/Qualifiers

1..419 /organism="Homo sapiens"
/map="14"

STS 1..150

primer_bind 1..22

primer_bind complement(127..150)

BASE COUNT 121 a 70 c 62 g 164 t 2 others

ORIGIN

Query Match 6.6%; Score 21; DB 5; Length 419;
Best Local Similarity 61.5%; Pred. No. 1.80e-03;

| Matches | 40; Conservative | 2; Mismatches | 22; Indels | 1; Gaps | 1; |
|-----------------------|--|--|------------|---------|----|
| Db | 46 ggaatgcacaactcatgagcgagcgggtttggtttatgtattgtttacaccaggatt | 105 | | | |
| Cp | 74 GGNNNNNNACCCMTGAGCGCGCGGT-GGTGTTACGTTGGTTTCTTGTAGGTTT | 16 | | | |
| Db | 106 aagaat 110 | | | | |
| Cp | 15 AGGAT 11 | | | | |
| RESULT | 11 | | | | |
| ID | HS1238538 | standard; RNA; EST; 421 BP. | | | |
| AC | AA437253; | | | | |
| NI | g2142167 | | | | |
| DT | 01-JUN-1997 (Rel. 52, Created) | | | | |
| DT | 01-JUN-1997 (Rel. 52, Last updated, Version 1) | | | | |
| DE | z54608.s1 Soares testis NHT Homo sapiens cDNA clone 757478 | 3' | | | |
| KW | EST. | | | | |
| OS | Homo sapiens (human) | | | | |
| OC | Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; | | | | |
| OC | Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| NC | [1] | | | | |
| RP | 1-421 | | | | |
| RA | Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S., | | | | |
| RA | Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J., | | | | |
| RA | Moore B., Schellenberg K., Steptoe M., Tan F., Theising B., | | | | |
| RA | White Y., Wylie T., Waterston R., Wilson R.; | | | | |
| RT | "WashU-Merck EST Project 1997"; | | | | |
| RL | Unpublished. | | | | |
| CC | Contact: Wilson RK WashU-Merck EST Project Washington University | | | | |
| CC | School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, | | | | |
| CC | MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: | | | | |
| CC | est@waton.wustl.edu This clone is available royalty-free through | | | | |
| CC | LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for | | | | |
| CC | further information. Seq primer: -41m13 fwd. ET from Amersham High | | | | |
| CC | quality sequence stop: 291. | | | | |
| FH | Key | Location/Qualifiers | | | |
| FH | 1..421 | | | | |
| FT | source | /organism="Homo sapiens" | | | |
| FT | | /notes-Vector: pT73D-Pac (Pharmacia) with a modified | | | |
| FT | | polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA | | | |
| FT | | was prepared from mRNA obtained from Clontech Laboratories | | | |
| FT | | Inc., and primed with a Not I - oligo(dT) primer [5' | | | |
| FT | | TGTTTACCATTCTGAAGTGGAGCGCGCCAAATTTTTTTTTTTT 3']. | | | |
| FT | | Double-stranded cDNA was ligated to Eco RI adaptors | | | |
| FT | | (Pharmacia), digested with Not I and cloned into the Not I | | | |
| FT | | and Eco RI sites of the modified pT73 vector. Library was | | | |
| FT | | through one round of normalization to Cot5, and was | | | |
| FT | | constructed by Bento Soares and M. Fatima Bonaldo." | | | |
| FT | | /clone="757478" | | | |
| FT | | /clone_lib="Soares testis NHT" | | | |
| FT | | /sex="male" | | | |
| FT | | /lab host="DH10B" | | | |
| FT | | complement(<1..>421) | | | |
| FT | mRNA | | | | |
| SQ | Sequence 421 BP; 75 A; 122 C; 151 G; 73 T; 0 other; | | | | |
| Query Match | 6.68; Score 21; DB 59; Length 421; | | | | |
| Best Local Similarity | 82.8%; Pred. No. 1.80e-03; | | | | |
| Matches | 24; Conservative | 1; Mismatches | 4; Indels | 0; Gaps | 0; |

Feb 23 10:57

/home/fuller/feb1098/US-08-836-075A-2.rag

3

```

FT /label= any amino acid
FT Misc difference 69
FT /label= Arg
FT /note= "amino acid in this position is designated X in
FT the specification, but codon usage shows that the
FT only possible amino acid at this pos. is Arg"
FT Misc difference 71
FT /label= Ala, Pro
FT Misc difference 95
FT /label= Gly
FT /note= "amino acid in this position is designated X in
FT the specification, but codon usage shows that the
FT only possible amino acid at this pos. is Gly"
PN W09613590-A2.
PD 09-MAY-1996.
PF 23-OCT-1995; E04155.
PR 21-OCT-1994; EP-870166.
PR 28-JUN-1995; EP-870076.
PA (INNO-) INNOGENETICS NV.
PI Maertens G, Stuyver L;
DR N-PSDB; T27937.
PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
PT - used to develop probes and primers for new sub:types and vaccines
PT to prevent and treat infection
PS Claim 25; Fig 3; 150pp; English.
CC The sequences R96526-R96578 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/EI, NS4 or NS5B regions of the genome.
CC This sequence represents amino acids 1-103 from the HCV type 1d isolate
CC BNL7.
CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/EI and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides R96424-R96524. The
CC sequences can also be used to synthesise probes and primers for the
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
SQ Sequence 109 AA;

Query Match 100.0%; Score 781; DB 19; Length 109;
Best Local Similarity 100.0%; Pred. No. 8.90e-62;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mstnpkprktrntnrpxxxxpgggqivggvylprgrprgxrtrktseersqprg 60
    |||||||
Qy 1 MSTNPKPRKTRNTNRPRXXXXPGGGQIVGGVYLLPRGRPRGXVRATRKTSERSQPRG 60

Db 61 rrgpikaxregrswagqypwplynecgcwaxwllsprgrpnwqp 109
    |||||||
Qy 61 RRQPIPKAXRKGESWAQPGYPWPPLYNCGCGWAXWLLSPRGSRPNWGP 109

RESULT 2
ID R96539 standard; peptide; 117 AA.
AC R96539;
DE Hepatitis C virus type 4k isolated BNL7 amino acids 1-117.
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;

```

Feb 23 10:57

/home/fuller/feb1098/US-08-836-075A-2.rag

4

```

KW PCR; primer; probe; antibody; infection.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT /note= "any amino acid"
PN W09613590-A2.
PD 09-MAY-1996.
PF 23-OCT-1995; E04155.
PR 21-OCT-1994; EP-870166.
PR 28-JUN-1995; EP-870076.
PA (INNO-) INNOGENETICS NV.
PI Maertens G, Stuyver L;
DR WPI; 96-251460/25.
DR N-PSDB; T27950.
PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
PT - used to develop probes and primers for new sub:types and vaccines
PT to prevent and treat infection
PS Claim 25; Fig 3; 150pp; English.
CC The sequences R96526-R96578 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/EI, NS4 or NS5B regions of the genome.
CC This sequence represents amino acids 1-117 from the HCV type 4k isolate
CC BNL7.
CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/EI and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides R96424-R96524. The
CC sequences can also be used to synthesise probes and primers for the
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
SQ Sequence 117 AA;

Query Match 99.5%; Score 777; DB 19; Length 117;
Best Local Similarity 91.7%; Pred. No. 2.12e-61;
Matches 100; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 1 mstnpkprktrntnrpxxxxpgggqivggvylprgrprgxrtrktseersqprg 60
    |||||||
Qy 1 MSTNPKPRKTRNTNRPRXXXXPGGGQIVGGVYLLPRGRPRGXVRATRKTSERSQPRG 60

Db 61 rrgpikaxregrswagqypwplynecgcwaxwllsprgrpnwqp 109
    |||||||
Qy 61 RRQPIPKAXRKGESWAQPGYPWPPLYNCGCGWAXWLLSPRGSRPNWGP 109

RESULT 3
ID R92976 standard; Protein; 191 AA.
AC R92976;
DT 02-OCT-1996 (first entry)
DE Hepatitis C virus isolate 26 core protein.
KW HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
    hepatitis.
OS Hepatitis C virus.
PN W09605315-A2.
PD 22-FEB-1996.
PF 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.

```

PI Bukh J, Miller RH, Purcell RH;
DR WPI: 96-139709/14.
DR N-PSDB; T16650.
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
used to determine HCV genotype and as vaccines against HCV infection
PS Claim 4; Page 214; 340pp; English.
CC R92936-R92987 are HCV core proteins derived from 52 different HCV
isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may
CC be used in the prevention of HCV infection.
SQ Sequence 191 AA;

Query Match 99.5%; Score 777; DB 17; Length 191;
Best Local Similarity 90.8%; Pred. No. 2.12e-61;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Db 1 mstnbpqkrtkntnrrpmdvkgggqivggvylprgrgxrgrvratrktsersqprg 60
Qy 1 MSTNPKPQKRTKNTNRRPXXXXPVGQVYLLPRGRGXRVATRKTSERSQPRG 60
Db 61 rrgpikarregswagqgypwplynecgagwllsprgrspwgp 109
Qy 61 RRQPIPKAXRXEGRSWAQGYPPWPLYGNEGGMAXWLLSPRGRSPNWGP 109

RESULT 4
ID R92973 standard; Protein; 191 AA.
AC R92973;
DT 02-OCT-1996 (first entry)
DE Hepatitis C virus isolate 28 core protein.
KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
OS Hepatitis C virus.
PN W09605315-A2.
PD 22-FEB-1996.
PF 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH) US SEC DEPT HEALTH.
PI Bukh J, Miller RH, Purcell RH;
DR WPI: 96-139709/14.
DR N-PSDB; T16647.
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
used to determine HCV genotype and as vaccines against HCV infection
PS Claim 4; Page 211-212; 340pp; English.
CC R92936-R92987 are HCV core proteins derived from 52 different HCV
isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may
CC be used in the prevention of HCV infection.
SQ Sequence 191 AA;

Query Match 99.5%; Score 777; DB 17; Length 191;
Best Local Similarity 90.8%; Pred. No. 2.12e-61;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Db 1 mstnbpqkrtkntnrrpmdvkgggqivggvylprgrgxrgrvratrktsersqprg 60

Qy 1 MSTNPKPQKRTKNTNRRPXXXXPVGQVYLLPRGRGXRVATRKTSERSQPRG 60
Db 61 rrgpikarregswagqgypwplynecgagwllsprgrspwgp 109
Qy 61 RRQPIPKAXRXEGRSWAQGYPPWPLYGNEGGMAXWLLSPRGRSPNWGP 109

RESULT 5
ID R67591 standard; Protein; 502 AA.
AC R67591;
DT 07-SEP-1995 (first entry)
DE Hepatitis C virus SR037-5' gene product.
KW Hepatitis C virus; HCV; non-A non-B; SR037-5'; treatment.
OS Hepatitis C virus.
PN J06319563-A.
PD 22-NOV-1994.
PF 13-MAY-1993; 147133.
PR 13-MAY-1993; JP-147133.
PA (IMMO) IMMUNO JAPAN KK.
DR WPI: 95-040318/06.
DR N-PSDB; Q79143.
PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the
PT treatment of hepatitis C
PS Claim 21; Page 36; 41pp; Japanese.
CC Q79143 is the hepatitis C virus (HCV) gene SR037-5' cDNA, it encodes
CC the protein described in R67591. Both the cDNA and protein can be
CC used in the treatment of HCV infection
SQ Sequence 502 AA;

Query Match 99.5%; Score 777; DB 12; Length 502;
Best Local Similarity 90.8%; Pred. No. 2.12e-61;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Db 1 mstnbpqkrtkntnrrpmdvkgggqivggvylprgrgxrgrvratrktsersqprg 60
Qy 1 MSTNPKPQKRTKNTNRRPXXXXPVGQVYLLPRGRGXRVATRKTSERSQPRG 60
Db 61 rrgpikarregswagqgypwplynecgagwllsprgrspwgp 109
Qy 61 RRQPIPKAXRXEGRSWAQGYPPWPLYGNEGGMAXWLLSPRGRSPNWGP 109

RESULT 6
ID R67589 standard; Protein; 502 AA.
AC R67589;
DT 07-SEP-1995 (first entry)
DE Hepatitis C virus YS117-5' gene product.
KW Hepatitis C virus; HCV; non-A non-B; YS117-5'; treatment.
OS Hepatitis C virus.
PN J06319563-A.
PD 22-NOV-1994.
PF 13-MAY-1993; 147133.
PR 13-MAY-1993; JP-147133.
PA (IMMO) IMMUNO JAPAN KK.
DR WPI: 95-040318/06.
DR N-PSDB; Q79141.
PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the
PT treatment of hepatitis C
PS Claim 19; Page 34; 41pp; Japanese.
CC Q79141 is the hepatitis C virus (HCV) gene YS117-3' cDNA, it encodes
CC the protein described in R67588. Both the cDNA and protein can be
CC used in the treatment of HCV infection
SQ Sequence 502 AA;

| | | | | | |
|-----------------------|---|---|---------------|-------------|---------|
| Query Match | 99.5%; | Score 777; | DB 12; | Length 502; | |
| Best Local Similarity | 90.8%; | Pred. No. 2.12e-61; | | | |
| Matches | 99; | Conservative 1; | Mismatches 9; | Indels 0; | Gaps 0; |
| Db | 1 | mstnbpqarktkntnrpqdvkfpqggqivgvyllprgrprvgvratkteesqprg | 60 | | |
| | | | | | |
| Qy | 1 | MSTNPKPQRTKNTNRPRXXXXXPGGGQIVGGVYLLPRGRGXGVRATKRTSESRQPRG | 60 | | |
| | | | | | |
| Db | 61 | rrqipkarrpegswaqgypwplynvegqgwagllsprgsrpswgp | 109 | | |
| | | | | | |
| Qy | 61 | RQPIPRAXXEGRSWAQGYPPWPLYNVEGCGWAXWLLSPGRSRPNWCP | 109 | | |
| | | | | | |
| RESULT 7 | | | | | |
| ID | R67588 | standard; Protein; 3011 AA. | | | |
| AC | R67588; | | | | |
| DT | 07-SEP-1995 | (first entry) | | | |
| DE | Hepatitis C virus gene product 'HC-G9. | | | | |
| KE | Hepatitis C virus; HCV; non-A non-B; HC-G9; treatment. | | | | |
| OS | Hepatitis C virus. | | | | |
| Key | Location/Qualifiers | | | | |
| FT | Misc_difference 2506..2565 | | | | |
| FT | /note= "corresponding codons ATA CGC CAC ATT AAC | | | | |
| FT | TCC GTG TGG CAA GAC CTT CTG GAG GAC AAT | | | | |
| FT | ACA ACA CCT ATA GAC ACT ACC ATC ATG GCA | | | | |
| FT | AAG AAT GAG GTC TTC" | | | | |
| PN | J06319563-A. | | | | |
| PD | 22-NOV-1994. | | | | |
| PF | 13-MAY-1993; 147133. | | | | |
| PR | 13-MAY-1993; JP-147133. | | | | |
| PA | (IMMO) IMMUNO JAPAN KK. | | | | |
| DR | WPI; 95-040318/06. | | | | |
| DR | N-PSDB; Q79140. | | | | |
| PT | A hepatitis C virus gene and oligo-nucleotide(s) - used for the | | | | |
| PT | treatment of hepatitis C | | | | |
| CC | Claim 18; Pages 21-33; 41pp; Japanese. | | | | |
| PS | Q79140 is the hepatitis C virus (HCV) gene HC-G9 cDNA, it encodes | | | | |
| CCC | the protein described in R67588. Both the cDNA and protein can be | | | | |
| CC | used in the treatment of HCV infection | | | | |
| SQ | Sequence 3011 AA; | | | | |

| | Query Match | 99.5% | Score 777; | DB 12; | Length 3011; |
|-----|---|--------|---------------------|--------|---------------------------------|
| | Best Local Similarity | 90.8%; | Pred. No. 2,12e-61; | | |
| | Matches | 99; | Conservative | 1; | Mismatches 9; Indels 0; Gaps 0; |
| Db | 1 metnbpqrktkntnrpqdvkfpgggqivgvyllprprvgvratrkteesqrg | 60 | | | |
| | | | | | |
| -Qy | 1 MSTNFKPQRKTNTNRREPXXXXXPGGGIVGVYLLPRRGRCGXGVRATRKTSERSQRG | 60 | | | |
| | | | | | |
| Db | 61 rrapipkarrpegrswagpywplyngnegcwaqwllsprgsrpswgp | 109 | | | |
| | | | | | |
| Qy | 61 RROP IPKAXRXEGRSWAOPGYTPMP LYGNCGCWAIXWLLSPRGSRPNGP | 109 | | | |
| | | | | | |

| RESULT | 8 |
|--------|---|
| ID | R92977 standard; Protein; 191 AA. |
| AC | R92977; |
| DT | 02-OCT-1996 (first entry) |
| DE | Hepatitis C virus isolate z7 core protein. |
| KW | HCV; HCV; E1: envelope 1; core protein; HCV genotyping; antibody; vaccine; hepatitis. |
| OS | Hepatitis C virus. |
| PN | W09605315-A2. |

22-FEB-1996.
15-AUG-1995; U10398.
15-AUG-1994; US-290665.
(USSH) US DEPT HEALTH & HUMAN SERVICES.
(USSH) US SEC DEPT HEALTH.
Bukh J, Miller RH, Purcell RH;
WPI: 96-139709/14.
N-PSDB; T16651.
DNA and amino acid sequence of HCV envelope 1 and core proteins -
used to determine HCV genotype and as vaccines against HCV infection
Claim 4; Page 214-215; 340pp; English.
R92936-R92987 are HCV core proteins derived from 52 different HCV
isolates. Isolated cDNA sequences are used for the prodn. of primers
useful for detecting the presence of HCV in a sample, the primers
are also useful for HCV genotyping. Proteins encoded by the cDNAs
can be used in vaccines for immunising against HCV infection. The
proteins may also be used to detect antibodies against HCV in serum,
saliva, lymphocytes or other mononuclear cells. The antibodies may
be used in the prevention of HCV infection.
Sequence 191 AA;
SQ

```

Query Match      99.2%; Score 775; DB 17; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.21e-61;
Matches 99; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db      1 mstnbpqrktkntarrpmdvkfpgggqivgvyllprgrplgvttrtkteersqprg 60
Qy      1 MSTNPKPQRKTKNTARRPXXXXXXXXXGGQIVGVYLLPRGRGPGXGVRAKTKTSERSQPRG 60

Db      61 rrpipkarrsegswagqypwplynngcgwagwllprgprswgpp 109
Ov      61 RRPIPKAXRGRSWAOPGYPPWPLYNNGCGWAXWLLSPGRSNNMP 109

```

| | | |
|--------|---|--|
| RESULT | 9 | |
| ID | R92937 standard; Protein; 191 AA. | |
| AC | R92937; | |
| DT | 02-OCT-1996 (first entry) | |
| DE | Hepatitis C virus isolate US11 core protein. | |
| KW | HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine; | |
| KW | hepatitis. | |
| OS | Hepatitis C virus. | |
| PN | W09605315-A2. | |
| PD | 22-FEB-1996. | |
| PF | 15-AUG-1995; U10398. | |
| PR | 15-AUG-1994; US-290665. | |
| PP | (USSH) US DEPT HEALTH & HUMAN SERVICES. | |
| PA | (USSH) US SEC DEPT HEALTH. | |
| PI | Bukh J, Miller RH, Purcell RH; | |
| DR | WPI: 96-139709/14. | |
| DR | N-PSDB; T16611. | |
| PT | DNA and amino acid sequence of HCV envelope 1 and core proteins - | |
| PT | used to determine HCV genotype and as vaccines against HCV infection | |
| PS | Claim 4; Page 181-182; 340pp; English. | |
| CC | R92936-R92987 are HCV core proteins derived from 52 different HCV | |
| CC | isolates. Isolated cDNA sequences are used for the prodn. of primers | |
| CC | useful for detecting the presence of HCV in a sample, the primers | |
| CC | are also useful for HCV genotyping. Proteins encoded by the cDNAs | |
| CC | can be used in vaccines for immunising against HCV infection. The | |
| CC | proteins may also be used to detect antibodies against HCV in serum, | |
| CC | saliva, lymphocytes or other mononuclear cells. The antibodies may | |
| CC | be used in the prevention of HCV infection. | |
| SQ | Sequence 191 AA; | |

| | | | | | | | |
|-----------------------|--------|--------------|---------------|------------|-----|--------|------|
| Query Match | 99.2%; | Score | 775; | DB | 17; | Length | 191; |
| Best Local Similarity | 89.9%; | Pred. | No. 3.27e-61; | | | | |
| Matches | 98; | Conservative | 2; | Mismatches | 9; | Indels | 0; |
| Gaps | 0; | | | | | | |

| | | | |
|----|----|---|-----|
| Db | 1 | msnbpqrkktntnrpqdvkfpqggqivgvyllprgrprlgvratrktsersqprg | 60 |
| | | | |
| Qy | 1 | MSTNPKPQKTKNTNRPPXXXXPGGQIVGVYLLPRRGRPXGVRATRKTSERSQPRG | 60 |
| | | | |
| Db | 61 | rrqipkarrpegdrtwacgypwplynecgcwagwllsprgsrpswp | 109 |
| | | | |
| Qy | 61 | RRQIPKARXREGSNAQGYWPWLYNCGCGHAWLLSPRGSRPNWCP | 109 |
| | | | |

| | |
|--------|---|
| RESULT | 13 |
| ID | R44010 standard; protein; 191 AA. |
| AC | R44010; |
| DT | 12-MAY-1994 (first entry) |
| DE | Hepatitis C Virus core protein. |
| KW | HCV; non-A, non-B hepatitis virus; NANBHV; immunodominant region; |
| KW | capsid protein; vaccine; antigen. |
| OS | Hepatitis C virus. |
| EH | Key Location/Qualifiers |
| FT | Peptide 1..45 |
| FT | /label= S42G |
| FT | /notes= "peptide 1" |
| FT | Peptide 38..81 |
| FT | /label= P42Y |
| FT | /note= "peptide 2" |
| FT | Peptide 77..115 |
| FT | /label= R40R |
| FT | /note= "peptide 3" |
| FT | Region 1..45 |
| FT | /label= immunodominant_region |
| FT | /note= "novel antigenic peptides contain sequences |
| FT | from this region" |
| FT | PN EP-569309-A. |
| PD | 10-NOV-1993. |
| PF | 06-MAY-1993; 420183. |
| PR | 06-MAY-1992; FR-005763. |
| PA | (DALB/) DALBON P. |
| PA | (INMR) BIO MERIEUX. |
| PI | Dalbou P, Jolivet M; |
| DI | WPI; 93-353383/45. |
| PT | New antigenic polypeptide(s) from hepatitis C virus - derived |
| PT | from N-terminal region of core protein, and related antibodies, |
| PT | useful in diagnosis, vaccination and treatment |
| PS | Claim 1; Fig 1; 24pp; French. |
| CC | Of peptides 1-3 (see Features Table), peptide 1 was found to be |
| CC | from the immunodominant region of HCV core protein. Further analysis |
| CC | of the N-terminal 45 amino acids of the HCV core protein showed that |
| CC | some epitopes are present in the first 21 amino acids, some in 22-45 |
| CC | and some at the junction of these two subregions. Preferred antigenic |
| CC | peptides are derived from S42G; they are useful for detecting HCV |
| CC | antibodies and in vaccines against HCV. |
| SQ | Sequence 191 AA; |

| | | | | |
|-----------------------|-------|---------------------|-------|---------------------------------|
| Query Match | 99.3% | Score 775, | DB 8; | Length 191; |
| Best Local Similarity | 89.9% | Pred. No. 3.27e-61; | | |
| Matches | 98; | Conservative | 2; | Mismatches 9; Indels 0; Gaps 0; |

| | | | |
|--------|--|---|-----|
| Db | 61 | rrcpkarpqrgrtwagpywpplymnegcwgawllsprqsrwqp | 109 |
| | | : : : | |
| Qy | 61 | RRQPIPKARXRCRSWAQPGYPMPFLYCNCGGCMAXLLSPGRSRNPMP | 109 |
| RESULT | 14 | | |
| ID | AC | R92938 standard; Protein; 191 AA. | |
| AD | IC | R92938; | |
| DT | 02-OCT-1996 | (first entry) | |
| DE | Hepatitis C virus isolate S14 | core protein. | |
| KW | HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine; hepatitis. | | |
| OS | Hepatitis C virus. | | |
| PN | W09605315-A2. | | |
| PD | 22-FEB-1996. | | |
| PF | 15-AUG-1995; J10398. | | |
| PR | 15-AUG-1994; US-290665. | | |
| PA | (USSH) US DEPT HEALTH & HUMAN SERVICES. | | |
| PA | (USSH) US SEC DEPT HEALTH. | | |
| PI | Bukh J, Miller RH, Purcell RH; | | |
| PI | WPI; 96-139709/14. | | |
| DR | N-FSDB; T16612. | | |
| PT | DNA and amino acid sequence of HCV envelope 1 and core proteins - | | |
| PT | used to determine HCV genotype and as vaccines against HCV infection | | |
| PS | Claim 4; Page 192; 340pp; English. | | |
| CC | R92936-R92987 are HCV core proteins derived from 52 different HCV | | |
| CC | isolates. Isolated cDNA sequences are used for the prodn. of primers | | |
| CC | useful for detecting the presence of HCV in a sample, the primers | | |
| CC | are also useful for HCV genotyping. Proteins encoded by the cDNAs | | |
| CC | can be used in vaccines for immunising against HCV infection. The | | |
| CC | proteins may also be used to detect antibodies against HCV in serum, | | |
| CC | saliva, lymphocytes or other mononuclear cells. The antibodies may | | |
| CC | be used in the prevention of HCV infection. | | |
| CC | Sequence 191 AA; | | |

| | | | | |
|-----------------------|------------------|---------------------|-----------|-------------|
| Query Match | 99.2%; | Score 775; | DB 17; | Length 191; |
| Best Local Similarity | 89.9%; | Pred. No. 3.27e-61; | | |
| Matches | 98: Conservative | 2: Mismatches | 9: Indels | 0: Gaps |

| | | | |
|----|----|--|-----|
| Db | 1 | mstnpqqrkktmrntrpqdqkfpgggqivagvylprgrprlgyratrktaesqrqg | 60 |
| Qy | 1 | MSNPKQEKTKRNNRPXXXXXPGGQIVGGVYLPRGRPRXGVRAKTKTSRSPQRG | 60 |
| Db | 61 | rrgpdkarprgegtwaqpgypwplynecgqagwllsprgsrswp | 109 |
| Qy | 61 | RROPIPKAKRGRSMAQPGYPMPLYNESCQGAAXLLSPRSRNPWG | 109 |

| | | |
|--------|--|--|
| RESULT | 15 | |
| ID | R79222 standard; Protein; 967 AA. | |
| AC | R79222; | |
| DT | 08-DEC-1995 (first entry) | |
| DE | pHCV141-encoded sequence. | |
| DE | pHCV141; amyloid precursor protein; APP; hepatitis C virus; HCV; EL; | |
| KW | E2; fusion protein; HEK-293; vaccine; vector; pHC/CW; | |
| KW | protein secretion; glycosylation. | |
| OS | Synthetic. | |
| PN | W09520664-A. | |
| PD | 03-AUG-1995. | |
| PF | 27-JAN-1995; U01087. | |
| PR | 28-JAN-1994; US-188281. | |
| PA | (ABBO) ABBOTT LAB. | |
| PI | Desai SM, Devare SG, Watanabe S, Yamaguchi J; | |

DR WPI; 95-275449/36.
PT New mammalian expression systems for HCV proteins - express fusion
PT proteins comprising amyloid precursor protein and HCV E1 and/or E2
PT protein
PS Disclosure; Page 59-62; 89pp; English.
CC RNA from serum or plasma of a chimpanzee infected with HCV was
CC converted to cDNA and PCR amplified using primers based on HCV
CC sequences. 7 Adjacent HCV DNA fragments were generated which
CC together encoded the HCV sequence given in R79232. Fragments
CC from 2 clones, pHCV141 and pHCV150 (see R79223), were combined to
CC generate pHCV176 (see R79221). This was used to construct
CC APP-HCV-E1+E2 fusion proteins that were glycosylated and secreted
CC from HEK-293 transfectants.
SQ Sequence 967 AA;

Query Match 99.2%; Score 775; DB 13; Length 967;
Best Local Similarity 89.9%; Pred. No. 3.27e-61;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnbpqrktkntnrpqdkvfpgggqivggvllprgrlgvratrkttsersqrg 60
|||||
Qy 1 MSTNPKQ8KTKNTNRPPXXXXXPGGGQIVGGVLLPRGRCYGVATRTKTSERSQPRG 60
|||||
Db 61 rrgqipkrrpgrtwagqypwplyngcqwagwllsprgsrpswqp 109
|||||
Qy 61 RRQPIPKAXRXEGRSWAQPGIPWPLYNGEGCGMAXWLLSPRGSRPNWGP 109
|||||

Search completed: Mon Feb 23 11:00:48 1998
Job time : 61 secs.